HOW DOES THE CHROMATIN REMODELER ATRX IDENTIFY ITS TARGETS IN THE GENOME?

Diu Thi Thanh Nguyen

Saint Anne's College

A thesis submitted for the degree of

Doctor of Philosophy

Of the

University of Oxford

MRC Molecular Haematology Unit
Weatherall Institute of Molecular Medicine
University of Oxford

Trinity Term 2014

44800 words

Acknowledgement

First and foremost, I would like to thank Richard Gibbons and Doug Higgs for the opportunity to work in such a stimulating and rewarding environment. Their critical thinking, intelligent direction and enthusiasm for science and for my work have inspired me. They have given me the confidence and the motivation to pursue this work. I would like to reserve a special thank to Richard for his teaching, his frequent encouragement throughout my degree, and his proof-reading the manuscript of this thesis.

This work would not have been possible without funding from the Marie Curie Initial Training Network and the Medical Research Council, for which I am extremely thankful.

I would like to acknowledge everyone in the Higgs and Gibbons lab for their considerable time and energy expended in teaching me techniques in the early stage of my degree. My great appreciation goes to Hsiao and Chris Fisher for their help with the supporting data in chapter 3 and 4; to Barbara for her help with the Western Blot analysis in chapter 5; and to Chris Babbs for his consultation of the telomere repeat cloning. I would also like to thank Nick Proudfoot, Nadina and Ben in the Proudfoot lab for teaching me the DIP technique.

Writing this thesis has been one of the most difficult things I have done. I could never imagine achieving it without brilliant suggestions and fruitful discussions with many members of the lab, especially Hsiao, David, Caroline, Danuta and Bryony. Thanks to David, Caroline, Gareth and Ricarda who have spared their precious time to read various chapters of this thesis for their helpful comments on the scientific contents and language. I have learnt a great lesson about how to use English articles.

My time in the lab has been unforgettable because of Ricarda and David – thank you for sharing experiment failure moments, for late evening discussions on many scientific topics and for sharing thoughts about career development.

Thank you Gareth for taking care of me, for bearing my frustration and for your understanding my hard time in the last few months.

Finally, I will be forever indebted to my family, Mom, Dad and Thuy for their unconditional love, support and concern for my well-being while I'm away from home. Thank you for all of the sacrifices you've made on my behalf. This thesis is dedicated to you.

Abstract

Diu Thi Thanh Nguyen Doctor of Philosophy

Saint Anne's College Trinity Term 2014

ATRX is a chromatin remodeling protein associated with X-linked Alpha-Thalassemia Mental Retardation syndrome and cancers that use the Alternative Lengthening of Telomere pathway. In the absence of ATRX there is a DNA damage response associated with telomeres and the expression of certain genes are perturbed. Recent findings (Law et al, 2010 Cell) have shown that ATRX is preferentially enriched at GC-rich tandem repeats in the genome. The mechanism for this localisation is unknown but may be related to the potential for these GC-rich tandem repeats to adopt non-B form DNA structures; ATRX has been shown to bind such structures (G4) *in vitro*. This study aims to understand the specific factors of the repeats that signal ATRX targeting.

To address the research questions, an experimental system was developed, in which known targets, the ψζ VNTR and telomere repeats, were inserted into an inducible ectopic gene in the 293T-Rex cell line by site-directed recombination.

ATRX was found to be enriched at the ectopic repeats compared to an endogenous negative control suggesting that it is recruited by the repeats independent of its original context. Furthermore, ATRX enrichment increased upon transcription of the ectopic gene, and this was dependent on the orientation of the repeat with the non-template strand being G-rich. Interestingly, when the repeat was transcribed, the distribution of ATRX across the repeats was asymmetrical with most ATRX binding downstream of the repeat. Moreover, there was a direct correlation between the repeat size and level of ATRX bound: the longer the repeat the higher the increase in ATRX enrichment.

To determine the signal for ATRX binding, assays were performed to look for features which reflected the distribution of ATRX including H3K9me3, RNA polII, G4, R loops and DNA supercoiling. R loops look to be a strong candidate for the signaling of ATRX binding.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADD	ATRX-DNMT3B-DNMT3L
AID	Activation-induced cytidine deaminase
ALC1	Amplified in Liver Cancer 1
ALS/FTD	Amyotrophic lateral sclerosis, frontotemporal dementia
ALT	Alternative Lengthening telomere
APE	Abasic endonuclease
ATMDS	Myelodysplastic syndrome (MDS) associated with thalassemia
AtNDX	At4g03090 gene
ATR-X	Alpha-Thalassemia Mental Retardation, X-linked syndrome
ATRX	ATRX protein
ATRXt	Truncated form of ATRX
ATS	Antisense
BAF	Brg1 Associated Factors
BAP	Brahma Associated Proteins
BER	Base excision repair
BLM	Bloom syndrome
BRG1	Brahma-related Gene 1 (ATPase subunit)
BRM	Brahma (ATPase subunit)
BSA	Bovine serum albumin
CBP/p300	CREBB-binding protein/EP300 binding protein
CD	Circular Dichroism
CDC13	Cell division control protein 13
CEB1	CEB1 minisatellite
CHD	Chromodomain-Helicase-DNA binding
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation-sequencing
ColE1	Colicin E1
COOLAIR	Set of long noncoding antisense transcripts produced at FLC
CoTC	Co-transcriptional cleavage
CSR	Class switching region
$dATRX_L$	*Drosophila* ATRX long isoform
$dATRX_S$	*Drosophila* ATRX short isoform

DAXX	Death-domain associated protein
DEK	DEK proto-oncogene
DExx	DEAD and DEAH motif related domain
DHS	DNase Hypersensitive Site
DHX9	DEAH box helicase 9
DIP	DNA immunoprecipitation
DIST	Rhomboid family 1 (RHBDF1) gene
DMEM	Dulbecco's modified Eagle's medium
DMR	Differentially methylated region
DMS	Dimethyl sulphate
DMSO	Dimethyl sulfoxide
DNMT	DNA methyltransferase
Dox	Doxycycline
DSB	Double strand break
dsDNA	Double-stranded DNA
elF4A	Elongation factor 4A
ENCODE	Encyclopedia of DNA Elements
ER	Estrogen receptor
ES cells	Embryonic stem cells
FACS	Flourescence-Activated Cell Sorting
FANCJ	Fanconi Anemia group J helicase
FBS	Fetal bovine serum
FLC	Flowering locus C
Flp	Flippase
FRET	Fluorescence resonance energy transfer
FRT	Flippase Recognition Target
GALpr	GAL1-10 promoter
Gamma-H2AX	Phosphorylated H2AX
GATA zinc finger	(A/T)GATA(A/G) sequence binding zinc finger
GFP	Green Fluorescence Protein
GNATs	Gcn5-related N-acetyltransferases
GQN1	G-quartet nuclease 1
H3K36	Lysine 36 of histone 3

H3K4me3	Trimethyl lysine 4 of histone 3
H3K79	Lysine 79 of histone 3
H3K9me3	Trimethylated lysine 9 of histone H3
H3S10	Serine 10 of histone 3
H3S10P	Phosphorylated Serine 10 of histone 3
H4K16ac	Acetylated lysine 16 of histone 4
HAT	Histone acetyltransferase
HBA1, HBA2	Haemoglobin, alpha 1/2
HbH	Haemoglobin H
HBM	Haemoglobin, mu
HEK293	Human embryonic kidney 293
HELICC	Helicase_C subdomain
HEP3B	Human hematopa cell line
HIRA	Histone cell cycle regulator
hnRNP	Heterogeneous nuclear ribonucleoproteins
HO	Hop gene, required for mating type switching
HP1	Heterochromatin protein 1
HSA	Helicase-SANT
Hygromycin R	Hygromycin B resistance
IgG	Immunoglobulin G
INO80	Inositol requiring 80
IP	Immunoprecipitation
ISWI	Imitation Switch
Kb	Kilobase pairs
kDa	Kilodaltons
Lac-Z	β-galactosidase
MAZ4	Sequence element containing 4 copies of the MAZ (MYC-associated zinc finger protein) binding sequence
MDS	myelodysplastic syndrome
MEFs	Mouse embryonic fibroblasts
mH2A	MacroH2A
MRN	Mre11-Rad50-Nbs1 complex
mRNA	Messenger RNA

MutS alpha	MSH2-MSH6 heterodimer
MYST	HAT family of 4 members MOZ, Ybf2, Sas2 and Tip60
NHEIII	Nuclease hypersensitive element III
NKFB	Nuclear factor kappa-light-chain-enhancer of activated B cells protein
NMR	Nuclear magnetic resonance spectroscopy
NoRC	nucleolar remodeling complex
NURD	nucleosome-remodeling and histone deacetylation
PARP	poly ADP ribose polymerase
PBAF	Polybromo-associated BAF
PBAP	Polybromo-associated BAP
PBS	Phosphate-buffered saline
pCMV	Cytomegalovirus immediate-early promter
PCR	Polymerase chain reaction
PDS	Pyridostatin
PHD	Plant Homeodomain
Pif1	PIF1 5'-to-3' DNA helicase
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid) sodium salt
PML	Promyelocytic leukemia
PML-NB	Promyelocytic leukemia-Nuclear body
PolII	RNA polymerase II
Poly(A)	Poly-adenine
POT1	Protection of telomere 1
pRNA	Promoter-associated RNA
pSV40	Simian virus 40 promoter
PTM	Post-translational modification
PVDF	Polyvinylidene difluoride
Q-PCR	Quantitative polymerase chain reaction
QLQ	Gln, Leu, Gln motif domain
RC	Replication-coupled
rDNA	Ribosomal RNA genes
RecA	ATP-dependent recombinase A
RecBCD	DNA helicase that has 3 components RecB, RecC and RecD

RecQ	3′ → 5′ DNA helicase
REV1	A member of the Y family of DNA polymerase
REZ	R-loop elongation zone
RI	Replication-independent
RIPA	Radioimmunoprecipitation assay buffer
RIZ	R-loop initiation zone
RNA polII	RNA polymerase II
RPA2	Replication protein A2
RSC	Remodel Structure of Chromatin
RTEL	Regulator of telomere elongation helicase
Rtel42 or (CCCAAT)42	Clone containing 71 unit of telomere repeat of the reverse orientation
Rtel71 or (CCCAAT)71	Clone containing 71 unit of telomere repeat of the reverse orientation
SANT	"Swi3, Ada2, N-Cor, and TFIIIB" domain
Sen1	Senataxin 1
SET domain	Su(var)3-9, Enhancer-of-zeste and Trithorax domain
SETX	Senataxin
SF1/2	Superfamily 1/2 of helicases
siRNA	Small interfering RNA
SLIDE	SANT-like domain
SSC	Saline sodium citrate buffer
ssDNA	Singel-stranded DNA
SUC2	beta-fructofuranosidase, required for growth on sucrose and raffinose
Suv39H1/H2	suppressor of variegation 3-9 homolog 1/2
SWI/SNF	SWItch/Sucrose NonFermentable
SYDH	Standford/Yale/USC/Harvard
TAE	Tris-acetate buffer
TAM	Transcription-associated mutagenesis
TAR	Transcription-associated recombination
TEBPβ	telomere-binding protein β
Tel42 or	Clone containing 71 unit of telomere repeat of the physiological

(TTAGGG)42	orientation
Tel71 or (TTAGGG)71	Clone containing 71 unit of telomere repeat of the physiological orientation
TERRA	Telomere repeat–containing RNA
Tet	tetracycline
TetO2	Two copies of Tetracycline operator sequence
THO/REX	Complex of (suppressors of the **t**ranscription defects of ***h****pr1*Δ mutants by **o**verexpression-**Tr**anscription and **Ex**port)
TLS	Translesion
TopI	Topoisomerase I
TopII	Topoisomerase II
TSS	Transcription start site
UTR	Untranslated region
VNTR	Variable number tandem repeat
WRN	Werner syndrome
XNP	*Drosophila* homolog of ATRX
Xrn	5'-3' Exoribonuclease 2
YAC	Yeast artificial chromosome

1 General introduction

Eukaryotic genomes are packaged with histone proteins in a compact structure, called chromatin. We now know that far from being a mere solution to the problem of packaging a huge amount of DNA in eukaryotic cells, chromatin actively participates in controlling all DNA-dependent activities through changes in its structure. Chromatin modification and chromatin remodeling are the two main mechanisms influencing the chromatin structure, making it more condensed or more open and accessible to regulatory factors. Chromatin remodeling is a process performed by a set of specialised complexes called chromatin remodelers.

One such chromatin remodeler, the ATR-X syndrome protein, is the subject of this DPhil project. The importance of this protein is indicated by the effect of germline mutations leading to the Alpha-thalassemia Mental Retardation Syndrome and somatic mutations associated with a variety of cancers. Recent advances in technology has led to some understanding of the structure and function of ATRX. However, the exact mechanism of its basic remodeling activity, in which finding remodeling targets is the first step, remains undetermined.

This chapter aims to introduce and summarise our knowledge of the topics that are related to this project. The first part of this chapter is a brief summary about chromatin structure. I then go on to discuss our current understanding of chromatin remodelers in general and the remodeler of interest ATRX, in particular. Finally, an introduction to non-canonical DNA structures will provide a background for the findings of this study.

1.1 CHROMATIN STRUCTURE

Genomic DNA of eukaryotes is packaged with protein histones in a compact structure, called chromatin. This is not only an excellent solution to the problem of packaging a huge amount of DNA in eukaryotic cells (e.g. ~ 6 billion bases in mammals), but also provides many layers of regulation for a variety of cellular activities such as DNA replication, transcription and repair (Khorasanizadeh 2004).

1.1.1 Nucleosome

Chromatin is composed of repeating units, called nucleosomes. Each nucleosome is formed by approximately a 146bp piece of DNA wrapped around a histone protein core particle, which consists of 2 copies of H2A, H2B, H3 and H4 protein. Nucleosomes are connected by a short DNA fragment called a linker, which varies in length from a few to 80bp, to form an array of nucleosomes. This organisation looks like a "beads-on-string" structure under electron microscopy (Luger, Dechassa et al. 2012).

The core particle of nucleosomes has a disc-shaped structure and is tightly wrapped by approximately 1.7 turns of left-handed superhelix (Khorasanizadeh 2004). All the histones that make up the octamer particle are small proteins (102-135 amino acid) and share a motif called the histone fold, through which histone components interact with each other to form a nucleosome (Bruce Alberts 2002). To assemble an octamer, the histone folds first associate with each other to form two heterodimers H3-H4 and H2A-H2B. Then two H3-H4 dimers combine together to form a tetramer, which further combines with two H2A-H2B dimers to make an octamer (Arents, Burlingame

et al. 1991). The nucleosomes are stabilised by a multitude of protein-protein and DNA-protein interactions. Most of the DNA-protein interactions occur within the core regions of the histone proteins, leaving their flexible N-terminal tails protruding out from the core (Margueron and Reinberg 2010). The histone tails are subject to various types of post-translational modifications, which will be discussed later on.

1.1.2 Higher order structures

In the presence of linker histone H1 or H5, the "beads-on-string" structure further folds into a chromatin fiber with a diameter of 30nm. There are two models explaining the formation of the chromatin fiber. Despite many attempts, it is still an unresolved question as to which model is correct. The first model corresponds to "one start" solenoid arrangement, in which consecutive nucleosomes interact with each other and follow a helical trajectory with bending of the linker DNA. The second model is called "two start" zigzag folding, which is an arrangement of two nucleosome rows with alternate nucleosomes interacting with each other and a relatively straight linker DNA. A recent study, however, suggests that there is no uniform configuration of the chromatin fiber, but rather a heteromorphic fiber with both zigzag and solenoid configurations. Subsequent fiber-fiber interaction contributes to the folding of chromatin into higher order structures (Li and Reinberg 2011).

1.1.3 Histone variants

It is now clear that apart from the general role of packaging DNA, protein histones also contribute to variations in chromatin structure to ensure dynamic patterns of gene

regulation throughout development. The incorporation of histone variants into chromatin plays a key role in this process. There are a growing number of histone variants that have been discovered. Except for histone H4, which has only one form identified so far, other core histones and linker histones have isoforms which differ from canonical histones by one to a few amino acids. Linker histone H1 family is the most diverse of all histones with 11 isoforms found in mice and 10 found in humans (Maze, Noh et al. 2014). Unlike core histones, histone variants are incorporated into chromatin independently from DNA replication throughout the cell cycle.

For the relevance of this thesis, I will discuss briefly here only the histone variants of the H2A and H3 families. Details about variants of other histone families are reviewed here (Maze, Noh et al. 2014).

1.1.3.1 H3 histone variant H3.3

Four different isoforms of H3 have been found: H3.1, H3.2, H3.3 and CENP-A. Among these, H3.3 and CENP-A have been studied the most intensively. H3.3 is different from H3.1 and H3.2 only in 4 or 5 amino acids (Sarma and Reinberg 2005). There are two H3.3 variants H3.3A and H3.3B encoded by *H3f3a* and *H3f3b* genes (Tang, Jacobs et al. 2013). H3.3 is deposited to various locations in the genome and its localisation changes during development, suggesting multiple roles of H3.3. H3.3 incorporation is both replication-coupled (RC) and replication-independent (RI), and it is carried out by several pathways depending on the sites of the localisation (Ahmad and Henikoff 2002) (De Koning, Corpet et al. 2007). RI deposition of H3.3 to active and repressed genes requires histone chaperone HIRA, whereas its localisation to

transcription factor binding sites is independent of HIRA (Goldberg, Banaszynski et al. 2010). Another important RI incorporation mechanism of H3.3 is through the chromatin remodeler ATRX and histone chaperone DAXX. The ATRX-DAXX complex deposits H3.3 to many heterochromatin regions of the genome, including telomeres, pericentric heterochromatin, silenced imprinted genes, and endogenous retroviral repeats (Goldberg, Banaszynski et al. 2010) (Voon, 2014, submitted).

1.1.3.2 H2A histone variant MacroH2A

For a long time, macroH2A (mH2A) has been known to interact with the inactive X chromosome (Xi) in female mammals (Costanzi and Pehrson 1998). However, recently it is also found in autosomes (Gamble, Frizzell et al. 2010). mH2A deposition onto Xi requires the localisation of the inactive-X-specific transcript, Xist (Mermoud, Costanzi et al. 1999). Chromatin immunoprecipitation and microarray data have shown a negative correlation between mH2A binding and gene expression, indicating the role of this histone variant in transcription repression. Interestingly, mH2A has been found to positively regulate transcription when located in the transcribed regions of a subset of its targets in autosomes (Gamble, Frizzell et al. 2010). This suggests that the roles of mH2A in gene regulation are diverse.

1.1.4 Histone modifications

The N-terminal tails of histones are protruding out from the core particle of nucleosomes and can be modified post-translationally. At least 8 types of post-translational modification (PTM) of histone tails have been discovered. Among these, small covalent modifications including acetylation, methylation and phosphorylation

are the most understood. Over 60 different histone residues have been found to be modified by specific antibodies or mass spectrometry (Kouzarides 2007). Histone modification involves modifying enzymes that add or remove post-translational modifications from amino acids on the core histones, recruiting other proteins and changing local chromatin organization (Petty and Pillus 2013) (Bannister and Kouzarides 2011, Langst 2013).

1.1.4.1 Histone methylation

Histone methyltransferases are responsible for adding methyl groups to lysine and arginine residues. All of histone methyltransferases that modify N-terminal lysines contain a catalytic SET domain. Methylation at lysines can happen at three levels: mono-, di- and trimethylation (Bannister and Kouzarides 2011). Lysine methylations that are associated with transcription activation are H3K4, H3K36 and H3K79. H3K4me3 and H3K36me3 have been implicated in transcription elongation. On the other hand, methylation at lysine 9, 27 and 40 of histone H3 are associated with transcription repression (Kouzarides 2007). H3K9me3 is the hallmark of heterochromatin including pericentric and telomeric regions. It is also enriched at promoters of silenced genes. However, H3K9me3 is also found in the body of active genes (Hahn, Wu et al. 2011). While constitutive heterochromatin is enriched with H3K9me3, euchromatin and facultative heterochromatin is marked primarily by H3K9me1/2 and limited regions of H3K9me3 (Collins and Cheng 2010). H3K9me2 is "written" by G9a, GLP, Eset/Setdb1 and Riz histone methyltransferases (Kubicek, O'Sullivan et al. 2007) and "erased" by JHDM2A, JMJD2 and LSD1 demethylases

(Collins and Cheng 2010). This histone modification has major functions in transcription control (Jenuwein 2006).

Arginines of histone H3 tails can be modified by type I and II arginine methyltransferases to one of three different forms mono-, symmetrically di- or asymmetrically di-methylated. Like lysine methylation, arginine methylation can be activatory or repressive to transcription (Kouzarides 2007).

1.1.4.2 Histone acetylation

Histone acetylation is highly dynamic and strongly linked to transcription activation (Verdone, Caserta et al. 2005). There are three families of histone acetyltransferases (HATs), GNAT, MYST and CBP/p300 (Kouzarides 2007). The transfer of an acetyl group from acetyl CoA to lysine, catalysed by the HATs, neutralise the positive charge of this residue, weakening the histone-DNA interaction. By doing so, acetylation makes chromatin more accessible to regulatory factors. Acetyl groups of lysines can be removed by histone deacetylases, and this event is correlated with transcription repression (Kouzarides 2007).

1.1.4.3 Histone phosphorylation

Histone phosphorylation occurs on serine, threonine and tyrosines residues, predominantly, but not exclusively in the N-terminal histone tails. This modification is "written" and "erased" by kinases and phosphatases, respectively (Oki, Aihara et al. 2007). All the known histone kinases catalyse the reaction transferring a phosphate group from an ATP to the hydroxyl group of the target amino acid side chain, adding

a substantial amount of negative charge to the histone. As a consequence, this certainly influences the chromatin structure (Bannister and Kouzarides 2011). Little is known about the function of this modification in gene regulation. H3S10 phosphorylation has been shown to activate NFKB-regulated genes and affects the binding of HP1 to H3K9me3 (Kouzarides 2007) (Fischle, Tseng et al. 2005).

1.2 CHROMATIN REMODELERS AND MECHANISM OF ACTION

1.2.1 Chromatin remodelers

As discussed in the previous part, DNA is packaged into chromatin. However, such embedding of DNA into chromatin presents a strong barrier to sequence specific recognition of DNA binding molecules, inhibiting all kinds of DNA dependent processes. To solve the problem of DNA accessibility, cells have developed two mechanisms: histone modification and chromatin remodeling. While a histone modification leads to recruitment of other proteins as well as direct changes in chromatin structure, the second mechanism involves a set of specialized molecules, called chromatin remodelers, to regulate accessibility to DNA. Chromatin remodelers utilize ATP hydrolysis to translocate, evict or assemble nucleosomes, hence exposing or protecting underlying DNA from regulatory factors (Clapier and Cairns 2009). Most, but not all, chromatin remodelers function in large macromolecular complexes (Petty and Pillus 2013). As exposure of chromatin to specialized effector proteins is vital to execute various activities including DNA replication, repair, homologous recombination and transcription, chromatin remodeling enzymes play a key role in virtually all of these processes, all of which require dynamic assembly, movement or displacement of nucleosomes.

1.2.2 Remodeler properties

Until recently, chromatin remodelers were classified into four families: SWI/SNF, ISWI, CHD and INO80. However, it is now accepted that ATRX, which used to be considered as a SWI/SNF family member, forms a new separate group of chromatin remodelers (Bartholomew 2014). All five groups share an ATPase domain, and are distinguished from each other by their unique flanking domains. The ATPase domain consists of two RecA-like boxes: DExx and HELICc, which is a common structural characterization of helicase-like proteins. Extensive comparison of protein sequences revealed that chromatin remodelers belong to the SNF2 family within the superfamily of helicase-like SF2 (Flaus, Martin et al. 2006). Chromatin remodeling enzymes are not *bona fide* helicases because they lack a wedge-like DNA-duplex destabilizing domain, which is a property of helicases in SF1 and SF2 superfamilies, enabling them to separate two nucleic acid strands (Saha, Wittmeyer et al. 2006). However, they exhibit DNA translocation activity that allow them to break DNA-histone contacts in order to reposition nucleosomes on a DNA molecule (Saha, Wittmeyer et al. 2002, Alexeev, Mazin et al. 2003, Jaskelioff, Van Komen et al. 2003, Whitehouse, Stockdale et al. 2003, Durr, Korner et al. 2005). Initial pieces of evidence for the ability of remodelers to translocate on DNA include the ability to displace the third strand in a DNA triple helix, and to redistribute or to assemble nucleosomes on DNA in an ATP-dependent manner (Saha, Wittmeyer et al. 2002) (Whitehouse, Stockdale et al. 2003, Xue, Gibbons et al. 2003).

1.2.3 Remodeler families

As mentioned above, five groups of chromatin remodelers are distinguished from each other by their flanking regions. The first way to catalogue remodelers is the linker region between DExx and HELICc boxes. SWI/SNF, ISWI, CHD and ATRX have a short linker while INO80 has a long linker. Each family is further defined by distinctive flanking domains. The SWI/SNF family contains a HSA domain and Bromodomain. ISWI and CHD families each have a SANT and a SLIDE domain, but CHD has a Plant homeodomain (PHD) and a Chromodomain that are absent in ISWI. ATRX possesses a unique ADD (ATRX-DNMT3-DNMT3L) domain, which consists of a GATA-like finger and a PHD finger. Although sharing the ability of changing the position of nucleosomes on DNA in an ATP-dependent manner, the outcome of remodeling varies depending on the type of remodelers, which will be discussed in the following section.

1.2.3.1 SWI/SNF family

The very first member of this family was discovered in two independent genetic screenings for mutations causing defects in expression of the HO gene, which is required for mating-type switching and in expression of the SUC2 gene needed for sucrose fermentation in yeast *S.cerevisiae*. It was revealed that these phenotypes were associated with mutations in SWI (Switching defective) gene and SNF (Sucrose Non Fermenting, SNF) gene, respectively (Neigeborn and Carlson 1984, Workman and Kingston 1998, Sudarsanam and Winston 2000). SWI/SNF genes encode a protein called ATPase Swi2/Snf2p, part of a larger complex which has 11 additional subunits. Purified SWI/SNF complexes were able to alter nucleosome positions in an ATP-

dependent manner (Vignali, Hassan et al. 2000). Another related member of SWI/SNF family also first described in yeast is the RSC complex composed of 17 subunits, in which Sth1 ATPase is a counterpart of the Swi2/Snf2p ATPase (Cairns, Kim et al. 1994).

The homologs of SWI/SNF and RSC complexes in *Drosophila* are BAP and PBAP which are built around the ATPase catalytic subunit BRM (Brahma), homologous to the yeast Swi2/Snf2 subunit. Similarly, in humans the two complexes are BAF and PBAF which is also comprised of the ATPase subunit BRG1 in their central core (Tang, Nogales et al. 2010). Other components of the two complexes in human and *Drosophila* also indicate homology to either yeast SWI/SNF or RSC complex.

The SWI/SNF family is defined by the presence of a N-terminal QLQ domain, probably important for protein-protein interaction (Kim, Choi et al. 2003), a HSA (Helicase-SANT) domain, which is known to recruit beta actin and actin related proteins (Szerlong, Hinata et al. 2008)and a C-terminal bromodomain, which is suggested to bind acetylated lysines of histone (Haynes, Dollard et al. 1992). A new, highly conserved domain, SnAC (Snf2 ATP Coupling) located between the HELICc box and bromodomain has recently been discovered to regulate ATPase and nucleosome mobilising activities of the SWI/SNF complex (Sen, Ghosh et al. 2011). This family of chromatin remodelers can perform nucleosome sliding and evicting activities but it lacks chromatin assembly activity (Clapier and Cairns 2009). Although SWI/SNF complexes were first described to have primary roles in transcription activation, evidence suggests that they also contribute to both gene repression and activation (Wilson and Roberts 2011). *Arabidopsis thaliana* SWI/SNF

complex has recently found to be associated with a RNA-mediated transcriptional silencing pathway via an interaction with long noncoding RNA (Zhu, Rowley et al. 2013).

1.2.3.2 ISWI family

Members of the ISWI (Imitation switch) family of chromatin remodelers are composed of 2 to 4 subunits (Clapier and Cairns 2009). The catalytic subunit of ISWI complexes in yeast is ISW1 and ISWI2, and in *Drosophila* ISWI. The mammalian homologues of ISWI are Snf2H and Snf2L. These core components are present in various combinations with auxiliary subunits, giving rise to different complexes with different properties (Langst 2013) . Well-characterised members of the family include yeast ISW1a, ISW1b and ISW2, human ACF/CHRAC and *Drosophila* NURF. Apart from the conserved ATPase subunit, ISWI remodelers also contain a SANT (Swi3 Ada2 N-CoR TFIIIB) domain and a juxtaposed SLIDE (SANT-like) domain at the C-terminal end (Boyer, Latek et al. 2004) (Grune, Brzeski et al. 2003) (Erdel and Rippe 2011). These domains are crucial for the binding of the ISWI remodeling complexes to nucleosomal DNA and for the recognition of histone tails (Boyer, Latek et al. 2004) (Grune, Brzeski et al. 2003). Extensive structural studies elegantly show that yeast ISW1a binds to two adjoining nucleosomes and sets the spacing between them (Yamada, Frouws et al. 2011).

ISWI remodeler functions are implicated in chromatin assembly and nucleosome spacing, which are essential in DNA replication, repair and transcription regulation (Erdel and Rippe 2011). An ISWI remodeler is a subunit of ACF, which was

demonstrated to establish an open chromatin structure downstream of the replication fork, crucial for replication progress through heterochromatin regions (Collins, Poot et al. 2002). Furthermore, Acf1, another component of the ACF complex, was shown to recruit Ku70/80 protein to DNA repair sites (Lan, Ui et al. 2010), and the related WSTF subunit of the WICH complex was reported to phosphorylate H2AX at DNA damage sites (Xiao, Li et al. 2009). One example of the transcription regulation role of ISWI remodelers can be taken from studies by the Langst group. They previously have shown that NoRC (Nucleolar remodeling complex) is recruited to silence ribosomal DNA genes and to establish heterochromatin formation (Strohner, Nemeth et al. 2001, Grummt and Langst 2013). In a recent study, with the aim to understand the mechanism of NoRC transcription silencing, they demonstrated that NoRC is able to specifically recognise and bind promoter nucleosomes with higher affinities than that for other nucleosomes (Manelyte, Strohner et al. 2014).

1.2.3.3 CHD family

The CHD (Chromodomain-Helicase-DNA binding) family of chromatin remodelers is characterised by the presence of two N-terminal tandem chromodomains, which mediate interactions with a variety of chromatin structures. The chromodomain in the CHD family is a conserved motif that shares homology with other chromodomains in HP1, Polycomb, histone acetyltransferases and histone methyltransferases (Paro and Hogness 1991, Jones, Cowell et al. 2000, Eissenberg 2001). Studies from several groups have suggested that the chromodomain facilitates the binding of CHD complexes with chromatin through directly interacting with DNA, RNA and methylated H3 (see review (Marfella and Imbalzano 2007)).

Based on additional motifs, the CHD family is further divided into 3 subfamilies: CHD1, Mi-2 and CHD7. The biological properties of the CHD remodelers are highly diverse. Yeast CHD1 has been shown to alter the structure of histone core particles in an ATP dependent manner distinct from SWI/SNF complexes (Tran, Steger et al. 2000). Chromatin assembly in a crude DEAE (Sepharose) extract of budding yeast is strongly affected in CHD1 deletion (Robinson and Schultz 2003). The study in *Drosophila* by Lusser and colleagues emphasised the roles of CHD1 in chromatin assembly. It demonstrated that CHD1 transfers histones from the NAP1 chaperone to DNA in a mechanism that gives rise to regularly spaced nucleosomes. Furthermore, CHD1 was shown to assemble H1-deficient chromatin but not H1-containing chromatin, suggesting its role in the assembly of active chromatin (Lusser, Urwin et al. 2005).

The functions of the CHD remodelers are highlighted in both transcription and its regulation. They are involved in every step of the transcription process including initiation, elongation, termination and RNA processing (see review (Murawska and Brehm 2011)). The NURD complex, containing histone deacetylases 1 and 2, is the only CHD remodeler that has been shown to be involved in transcription repression. It was proposed that NURD is recruited to the promoters of target genes where it remodels adjacent nucleosomes, facilitating deacetylation of their tails (Murawska and Brehm 2011).

1.2.3.4 INO80 family

The defining feature of INO80 (Inositol requiring 80) family is a long linker between the DExx and HELICc boxes of the ATPase domain, which is responsible for the association with RuvB-like proteins. RuvB is a microbial ATP-dependent helicase that catalyses branch migration of Holliday junctions during bacterial recombination. Thus, the presence of RuvB-related components in INO80 complexes make them the only chromatin remodeler family that exhibits helicase activity (Laura Manelyte and Gernot Längst, 2013). Similar to SWI/SNF family, the INO80 family also contains an HSA domain which is required for the recruitment of actin and actin-related proteins (Gerhold and Gasser 2014).

Structural analyses have given a fascinating insight into the way that INO80-C and SWR-C complexes, two members of the family, interact with their nucleosomal substrate. According to these studies, INO80-C grasps nucleosomes “like a hand” and SWR-C appears to cling to its substrate through the Swr1 component (Nguyen, Ranjan et al. 2013, Tosi, Haas et al. 2013) (Gerhold and Gasser 2014).

The INO80 remodelers display versatile remodeling activities. Purified yeast INO80 remodels chromatin and facilitates transcription *in vitro*. Mutations in this complex causes sensitivity to DNA damage agents and defects in transcription (Shen, Mizuguchi et al. 2000). Elegant work by van Attikum and Tsukuda links the roles of INO80 and SWR1 to DNA repair. The complexes were reported to be recruited to double strand break sites and evict H2A.Z, gamma H2A.X and core histones near the break (Tsukuda, Fleming et al. 2005, van Attikum, Fritsch et al. 2007) (Horigome, Oma et al. 2014). Udugama *et al*. observed that INO80 exhibited a nucleosome spacing activity. The study also compared the nucleosome mobilisation mode of this

complex to that of SWI/SNF and ISWI complexes (Udugama, Sabri et al. 2011). Furthermore, global localisation of H2A.Z is affected in the absence of INO80. It was also demonstrated that the remodeler can replace nucleosomal H2A.Z/H2B with free H2A/H2B dimers (Papamichos-Chronakis, Watanabe et al. 2011).

1.2.3.5 ATRX family

The feature that distinguishes the ATRX family from other families of chromatin remodelers is the ADD domain at the N-terminus. It also possesses an HP1 binding domain. Details about the structure and remodeling functions of this remodeler will be discussed in the following section.

1.2.4 Targeting mechanism of chromatin remodelers

Chromatin remodelers are highly diverse and abundant in cells. Their diversity is owed to numerous combinations of core catalytic domains and accessory components that form different remodeling complexes with distinct biological functions. For example, a yeast RSC complex has 15 subunits, and CHD and INO80 family members are usually composed of 1-10 subunits or even more (Cairns, Lorch et al. 1996). It is estimated that there are several hundred chromatin remodeling complexes in human cells (Manelyte, Strohner et al. 2014). In addition to the high diversity, chromatin remodeling enzymes are surprisingly abundant, given that one enzyme can successively catalyse multiple nucleosomes (Varga-Weisz 2010). Estimation by several groups based on remodeler and nucleosome concentrations in yeast and human cells indicates that approximately one enzyme is responsible for remodeling

10 nucleosomes (Huh, Falvo et al. 2003, Weidemann, Wachsmuth et al. 2003) (Ghaemmaghami, Huh et al. 2003) (Erdel, Schubert et al. 2010).

Remodeling complexes preferentially localise to specific genomic sites. This raises two questions, which signals target these complexes to these sites, and how do they search for the signals. The following sections discuss our current understanding of the mechanisms by which chromatin remodelers recognise their targets in the genome.

1.2.5 Targeting signals

1.2.5.1 DNA and RNA sequence and structure

The first indication that DNA sequences can be used as a targeting signal for chromatin remodelers lies in the fact that many remodelers possess a DNA-binding domain that is present in the catalytic and/or accessory subunits: for example, the SANT and/or SLIDE domains that are present in ISWI and SWI/SNF families, and the WAC and AT hook motifs in the ACF1 and TIP5 proteins.

Nucleosome occupancy can be predicted by the primary DNA sequences, which are associated with an ability to bend in a manner required for nucleosome formation (Partensky and Narlikar 2009). For this reason, the underlying DNA sequences can play important roles in directing chromatin remodelers to their nucleosomal substrates and in dictating the outcome of remodeling. Using single molecule imaging and high resolution native poly-acrylamide gel electrophoresis, van Vugt *et al*. reported that fly Mi-2 (CHD type) and yeast RSC (SWI/SNF type) complexes bind to nucleosomal positioning sequences, which strongly influence initial translocation direction, processivity and final octamer positioning (van Vugt, de Jager et al. 2009). In

addition, it was observed that the enzymes that relocate nucleosomes to more central locations, like ISW1a, ISW1b and CHD1 preferentially bind to nucleosomes bearing linker DNA. That is why these enzymes are unable to move nucleosomes close to DNA ends because they have lost the DNA linker (Stockdale, Flaus et al. 2006). However, a study by Partensky and Narlikar using FRET suggests that the rate of remodeling by human ACF and yeast RSC is independent of the affinity of the DNA sequence for histone octamers. This study also showed that the outcome of positioning is determined by sequence at the local level (Partensky and Narlikar 2009).

Methylated DNA can serve as a targeting signal for remodelers containing MDB (methyl-binding domain) domains. For example, the recruitment of the NURD complex to methylated CpG island promoters is mediated by MBD2 (Zhang, Ng et al. 1999).

In addition to the primary sequence, remodeling machines can also interpret special conformational features of DNA to establish different nucleosome positioning patterns. In particular, remodeling by ACF is directed by the presence of a DNA element that contains an intrinsically curved region (Rippe, Schrader et al. 2007). In addition, using the 3C technique Yadon *et al*. have recently presented evidence that ISW2 can be targeted to specific loci by transcription factor Ume6- and TFIIB-dependent DNA looping (Yadon, Singh et al. 2013). Furthermore, highly GC-rich DNA sequences, like telomeres or rDNA, have been shown to be enriched with the chromatin remodeler ATRX. Interestingly, these have been reported to form a four-stranded DNA structure, called G4, which binds ATRX *in vitro*. It is tempting to

speculate that the cells use this structure for remodeler localization (Law, Lower et al. 2010), however this remains to be elucidated.

Studies have shown that RNA structures also play an important role in the targeting mechanisms of chromatin remodelers. Manelyte *et al*. demonstrated that a non-coding RNA (pRNA), which is initiated upstream and contains the promoter sequence of rDNA genes, is required to recruit NoRC to silence transcription. It was found that the Tip5 domain of NoRC recognised the stem loop of the pRNA because the inhibition of remodeling by the NoRC complex was lost when the stem loop is mutated (Manelyte, Strohner et al. 2014).

1.2.5.2 Nucleosome occupancy

Evidence that targeting sites for remodelers can be marked by nucleosome depletion comes from the laboratory of Ahmad. His group previously showed that XNP, a *Drosophila* analog of the human remodeler ATRX, targeted dynamic chromatin and sites of rapid nucleosome replacement (Schneiderman, Sakai et al. 2009). They further demonstrated that the XNP remodeler, together with the histone chaperone, HIRA, recognised and bound to exposed DNA of active transcribed genes and was only displaced when new nucleosomes were assembled (Schneiderman, Orsi et al. 2012). It is unclear as to whether human ATRX is also recruited to nucleosome-free regions.

1.2.5.3 Histone modifications

As discussed in the previous section, residues on histone tails can be post-translationally modified. Post-translational histone modifications (PTMs) can be "read" by specific modules of effectors that facilitate downstream events (Taverna, Li et al. 2007). Chromatin remodeling complexes contain domains that can recognise PTMs at specific genomic locations, allowing establishment of remodeler dependent nucleosome positioning.

Acetylated histones are recognised by bromodomains, that are present in RSC, CHD1, ISWI and ACF1 complexes, which results in promotion or inhibition of remodeling activities (Erdel, Krug et al. 2011). For example, acetylated H3, but not H4, increase the bromodomain dependent recruitment of the SWI/SNF and RSC complexes to nucleosomes. The recruitment, however, is not further enhanced by additional bromodomains found in RSC (Chatterjee, Sinha et al. 2011).

With regards to histone methylation, the PHD (Plant homeodomain) domain of the NURF complex (ISWI type) has been shown to mediate a direct interaction with the H3K4me3 histone mark. It was reported that depletion of this modification led to partial disassociation of BPTF, a NURF subunit, from chromatin and impair the recruitment of the ATPase subunit SNF2L (Wysocka, Swigut et al. 2006). Elegant work by Eustermann and Iwase has shown that unmodified H3K4 and H3K9me3 directly interact with the ADD domain of ATRX, which contains a PHD finger and a GATA-like finger, to recruit ATRX to pericentric and telomeric heterochromatin (Eustermann, Yang et al. 2011) (Iwase, Xiang et al. 2011). Double chromodomains of the human CHD1 complex specifically associate with H3K4me2/3 but not unmodified

H3K4. However, a recent study showed that this interaction was not required for overall recruitment of CHD1 to chromatin *in vivo* (Morettini, Tribus et al. 2011).

1.2.5.4 Histone variants

Chromatin remodelers are responsible for replication-independent incorporation of the histone variants H3.3 (CHD1, ATRX) (Morettini, Tribus et al. 2011) (Goldberg, Banaszynski et al. 2010), H2A.Z (ISWI) (Goldman, Garlick et al. 2010) and CENH3/CENP-A (CHD1 and RSF) (Okada, Okawa et al. 2009) (Perpelescu, Nozaki et al. 2009). In some cases histone variants represent the products of the remodeling process whereas in other cases, they play the role of a targeting signal although it can be hard to distinguish between these two. Although H3.3 and ATRX are co-immunoprecipitated, it is now clear that ATRX is specifically required for localisation of H3.3 to telomeres (Goldberg, Banaszynski et al. 2010).

When Goldman *et al*. purified chromatin from cells expressing epitope-tagged H2A or H2A.Z and looked for enrichment of remodeler proteins, they found that all major classes of chromatin remodelers were differentially associated with H2A.Z. Furthermore, H2A.Z incorporation was shown to increase the nucleosome remodeling activity of human ISWI (Goldman, Garlick et al. 2010).

Another example of the roles of histone variants in remodeler targeting refers to nucleosomes containing gamma H2A.X, which is important in DNA damage response. The SWI/SNF complex was found to directly bind gamma H2A.X containing nucleosomes through an interaction between its catalytic subunit, BRG1,

and H3 acetylation which is induced by the phosphorylation of H2A.X in response to DNA damage. The binding, in turn, facilitates the phosphorylation of H2A.X at Ser-139. This exemplifies the intricate feedback activation loop by chromatin remodeling machines, histone variants and histone modifications to facilitate DNA repair (Lee, Park et al. 2010). In this case it is difficult to differentiate the signaling role from the remodeling consequence of histone variants and modifications.

1.2.5.5 Recruitment factors

Chromatin remodelers can be recruited to chromatin through interaction with DNA-binding transcription factors or specialised effectors. The *Drosophila* transcription factor Tramtrack69 (TTK69) is able to bind chromatin in the absence of the NURD remodeling complex, but targeting of NURD is dependent on TTK69, suggesting transcription factor binding precedes remodeler recruitment. More specifically, TTK69 recruits NURD via interacting with MEP1, a subunit of NURD. Remarkably, transcriptome profiling showed that TTK69 correlated poorly with remodelers other than NURD, indicating the selectivity of this targeting pathway (Reddy, Bajpe et al. 2010).

Similarly, Sox10 has been shown to recruit BRG1 containing complexes via physically interacting with BAF60, a facultative subunit of the complexes, to two key target genes in Schwann cells (Weider, Kuspert et al. 2012). Recruitment of SWI/SNF to target genes of the estrogen receptor α (ER α) requires the nuclear receptor coactivator protein Flightless-I (Fli-I) binding to BAF53, an actin-related component of the SWI/SNF complex (Jeong, Lee et al. 2009).

Several studies have provided evidence for the role of poly(ADP-ribose) in localising CHD4 chromatin remodelers to DNA double-strand breaks (Polo, Kaidi et al. 2010). The ALC1 (Amplified in Liver Cancer 1), also known as CHD1L, has been identified as a novel chromatin remodeling enzyme that is rapidly recruited to DNA damage sites in a poly(ADP-ribose) dependent manner to catalyse nucleosome sliding (Ahel, Horejsi et al. 2009) .

1.2.6 Substrate search mechanism

In order to identify their target sites in the genome, chromatin remodelers have to detect signals that mark these sites. Studies have proposed a number of genomic screening mechanisms by the remodelers. Perhaps the most attractive searching model is the continuous sampling/immobilisation mechanism, postulated by the Rippe's laboratory. According to this mechanism, remodeling complexes constantly, but transiently, sample nucleosomes without causing any remodeling effect. Only upon recognition of a specific targeting signal, which is discussed above, is a stable interaction with a remodeler established (Varga-Weisz 2010). With the focus on the ISWI class, they used imaging of fluorescently tagged chromatin remodelers to study their dynamics in living human and mouse cells. They discovered that SNF2H and SNF2L are relatively immobile at DNA replication and repair foci, whereas they are highly mobile during G1 and G2 phases with a diffusion rate incompatible with the time required for a remodeling reaction as measured by single-molecule analysis (Blosser, Yang et al. 2009). These data are consistent with the scenario in which ISWI complexes are continuously "moving and hitting" chromatin unless a specific feature

causes them to engage with the chromatin at a specific site (Erdel, Schubert et al. 2010).

Another substrate searching mechanism, termed the release/termination model, that has been discussed in the literature involves the Michaelis Menten enzyme kinetics mechanism. In this screening model, chromatin remodelers play the role of an enzyme, and the substrates are nucleosomes, which follow Michaelis Menten-like equation. The "good" substrates are the nucleosomes that have a high affinity to the chromatin remodeling enzyme (low value of the Michaelis-Menten constant K_M) and a high catalytic conversion rate of the enzyme-substrate complex to the product (repositioned nucleosomes) in a translocation reaction (high value of the catalysis constant K_{cat}). An efficient catalytic process is characterised by a high ratio of K_{cat}/K_M, and a low ratio represents a "bad" nucleosome substrate. According to the release model, a remodeler binds to good substrates and converts them to bad substrates (products of remodeling), which have low affinity to the remodeler and hence release it. Alternatively, a remodeler recognises all of its substrates with similar affinities, but it has a slow translocation rate on "bad" substrates (Rippe, Schrader et al. 2007) (Manelyte and Langst, 2013).

1.2.7 Remodeling mechanism

Studies of remodeling mechanisms have been addressing two questions: (1) how do chromatin remodelers use ATPase-dependent translocase activity to disrupt histone-DNA contacts and (2) how do different remodelers apply this property to generate the different outcomes of remodeling.

The ATPase domain of chromatin remodelers is similar to known DNA translocases, and has been shown to perform DNA translocation by a large number of studies (Clapier and Cairns 2009). Single-molecule approaches have been used to detect the ability of SWI/SNF complexes to create ATP-dependent loops on nucleosomal DNA molecules (Fitzgerald, DeLuca et al. 2004) (Saha, Wittmeyer et al. 2006) (Strohner, Wachsmuth et al. 2005, Zhang, Smith et al. 2006) (Lia, Praly et al. 2006). This translocation can happen in both directions 3'->5' and 5'->3' with a speed of 25bp per second and is likely to be made of 2bp steps. More recently, repositioning of ISWI complexes has been shown to occur in 1bp steps (Bartholomew 2014). When an enzyme translocates along DNA, it creates an accumulation of superhelical torsion. It is thought that the combined effect of translocation and torsion could create incremental distortions of histone-DNA contacts which are harnessed by different enzymes to generate different outcomes (Narlikar, Sundaramoorthy et al. 2013). In support of this model, translocase domains of chromatin remodelers have been shown to engage DNA at an internal region of a nucleosome (Narlikar, Sundaramoorthy et al. 2013). Regarding the histone octamer, recent data suggest that nucleosomes are able to adopt conformation changes at the interface between H2A-H2B and H3-H4 tetramer. This presumably increases the types of possible DNA distortion within a nucleosome (Bohm, Hieb et al. 2011).

Another mechanistic question is as to how the ATPase activity of remodelers is regulated. Studies have suggested that the regulation can be achieved through histone

modifications and the linker DNA. For example, ATPase has been shown to be stimulated by a region on the H4 tail (residue 17-19) and attenuated by H4K16ac (Clapier, Langst et al. 2001, Hamiche, Kang et al. 2001, Clapier, Nightingale et al. 2002). Furthermore, ISWI remodelers are regulated by the length of DNA linker of the nucleosome, which is important during chromatin assembly (Yang, Madrid et al. 2006).

1.3 THE CHROMATIN REMODELER ATRX

1.3.1 Structure of the ATRX gene and protein

The human ATRX (Alpha Thalassemia/mental Retardation syndrome X-linked) gene is comprised of 35 exons and spans approximately 300kb on chromosome Xq13. The ATRX protein consists of two known domains, an ATRX-DNMT3-DNMT3L (ADD) domain at the N terminus and a helicase/ATPase at the C terminus. The ATRX gene gives rise to two major protein isoforms: a full length 280kDa protein and a truncated form (ATRXt) of 180kDa, which arises due to an alternative splicing event at intron 11 and the use of a proximal intronic poly-A signal and hence it does not contain the ATPase domain (Garrick, Samara et al. 2004).

The cysteine-rich ADD domain is comprised of three distinguishable modules, a GATA-like zinc finger and a PHD zinc finger at the N terminus and a long C-terminal α helix (Argentaro, Yang et al. 2007). The ATRX PHD finger is similar to that of DNMT3A, DNMT3B and DNMT3L in the way that it is flanked by an N-terminal C_2C_2 motif. PHD zinc fingers are present in nuclear proteins and have a role in tethering proteins to chromatin (Bienz 2006). The extended C-terminal helix makes an

extensive hydrophobic contact with the GATA-like finger, bringing the N and C termini together (Argentaro, Yang et al. 2007). This domain organisation is of great importance in the ATRX function in recognising specific histone modifications, which will be discussed further later in this section.

In addition, the human ATRX also contains an HP1 interaction motif which is located downstream of the ADD domain (Eustermann, Yang et al. 2011). This motif Leu-X-Val-X-Leu of ATRX has been shown to be responsible for interaction with the chromo shadow domain of HP1α and HP1β (McDowell, Gibbons et al. 1999) (Thiru, Nietlispach et al. 2004) (Lechner, Schultz et al. 2005).

Orthologs of the human ATRX protein have been found in other organisms. It is observed that the ATPase domain is highly conserved through evolution. The mouse ATRX is almost identical to the human protein at the two highly conserved domains ADD and ATPase (Picketts, Tastan et al. 1998). The putative homolog of ATRX in fly, called XNP/ATRX, is present in two isoforms, $dATRX_S$ and $dATRX_L$ that are derived from the use of two alternative translational start codons (Bassett, Cooper et al. 2008). Both dATRX isoforms have the conserved ATPase domain but lack the ADD domain (Valadez-Graham, Yoshioka et al. 2012). The nematode (*C. elegans*) gene of homologous ATRX protein, termed *xnp-1*, also lacks sequences coded by the first nine exons of the human gene where the ADD domain is located (Villard, Fontes et al. 1999).

1.3.2 ATRX as a chromatin remodeler

1.3.2.1 Remodeling and translocation activity of ATRX

ATRX has been shown to be able to remodel nucleosomes in a different mode, compared to SWI/SNF remodelers. To demonstrate this, Xue *et al*. employed a rotationally phased mononucleosome, prepared using sea urchin 5S rDNA containing a nucleosome positioning sequence, to test if ATRX complex could alter the DNaseI digestion pattern, which was characterised as a distinctive 10bp ladder on a PAGE gel. It was observed that the complex altered the digestion pattern at the region near the entry site of the nucleosome in the ATP-dependent manner. However, unlike SWI/SNF, it did not disrupt the nucleosome phasing, suggesting that ATRX complex remodels chromatin in a different manner (Xue, Gibbons et al. 2003).

Furthermore, ATRX-DAXX complexes are able to displace the third strand in the triplex helix displacement assay, but they do not have a DNA helicase property. It's worth noting that the assay was performed with the presence of DAXX in the ATRX complexes, however DAXX is dispensible for the translocation activity of ATRX (Xue, Gibbons et al. 2003).

In addition, ATRX has been tested for G-quadruplex (G4) unwinding activity. Previously it has been reported to bind G4 *in vitro* (Law, Lower et al. 2010); however unlike BLM (Bloom syndrome) helicase it does not unwind G4 DNA (Clynes, Jelinska et al. 2014).

Taken together, these *in vitro* studies suggest that ATRX is indeed a chromatin remodeler, not a helicase, and it remodels chromatin in a different way than SWI/SNF

complexes. The combined differences in structure (discussed above) and remodeling activities, compared to SWI/SNF family, have classified ATRX as a separate remodeler class (Bartholomew 2014). So what have we learned about the remodeling functions of ATRX *in vivo*? What are specific types of remodeling that ATRX performs in the cells?

1.3.2.2 Deposition of H3.3 histone variant at pericentric and telomeric heterochromatin

This topic has been a fascinating area of research for the past few years. It is becoming clear that the temporal and spatial locations of ATRX-dependent deposition of H3.3 are diverse, reflecting the multiple roles ATRX plays in chromatin dynamics. Two genomic regions that has been extensively described into which H3.3 is deposited by ATRX: pericentric heterochromatin and telomeres (Wong, McGhie et al. 2010) (Goldberg, Banaszynski et al. 2010) (Lewis, Elsaesser et al. 2010, Ivanauskiene, Delbarre et al. 2014) (Drane, Ouararhni et al. 2010).

It is now clear that the general mechanism of H3.3 localisation at both regions follows the axis ATRX-DAXX-H3.3. DAXX (Death associated protein) is a histone chaperone that has been shown to envelop H3.3-H4 dimers for H3.3 recognition (Drane, Ouararhni et al. 2010) (Elsasser, Huang et al. 2012) (Delbarre, Ivanauskiene et al. 2013). ATRX plays a guiding role: it localises at these regions and directs the deposition. Knocking-out of either ATRX or DAXX leads to loss of H3.3 at telomeres in mouse embryonic stem (ES) cells. Similar levels of reduction are observed in both cases (Lewis, Elsaesser et al. 2010) (Goldberg, Banaszynski et al. 2010). Likewise, depletion of ATRX (by siRNA) or DAXX reduces H3.3 enrichment at pericentric

heterochromatin in mouse embryonic fibroblasts (MEFs) (Drane, Ouararhni et al. 2010). These suggest that ATRX and DAXX work in the same pathway. Interestingly, in the absence of ATRX, DAXX fails to localise to telomeres, suggesting that ATRX is required to target DAXX and H3.3 to these regions. Furthermore, ATRX levels are decreased in DAXX-deficient cells, indicating that DAXX might be important for ATRX stability and expression (Lewis, Elsaesser et al. 2010).

In mouse embryonic stem (ES) cells, ATRX is recruited to telomeres during S phase of the cell cycle. There is strong evidence that ATRX deficiency affects telomere integrity, suggesting the role of ATRX-dependent H3.3 deposition in maintaining telomere stability (Wong, McGhie et al. 2010). This raises a question of the underlying mechanism. On the other hand, ATRX has been shown to deposit H3.3 outside S phase in MEFs and knocking down ATRX, DAXX or H3.3 results in a decrease in pericentric transcripts, indicating that ATRX/DAXX recruitment of H3.3 plays a role in regulating transcription of these repeats (Drane, Ouararhni et al. 2010). On the contrary, loss of ATRX in mouse ES cells leads to an increase in telomeric RNA repeat containing RNA (TERRA) (Goldberg, Banaszynski et al. 2010). It is possible that ATRX regulates transcription of TERRA and pericentric repeats in different mechanisms.

A recent study in our lab has found that that ATRX-dependent deposition of H3.3 is not only limited to pericentric heterochromatin and telomeres, but also takes place at other interstitial heterochromatic repeat regions throughout the genome, including tandem repeats, endogenous retroviral repeats and non-promoter CpG islands. These intragenic CpG islands fall into a class of genomic sites called imprinted control

regions which are characterised by allelically differential DNA methylation (imprinted DMRs), a high level of H3.3 histone variant and H3K9me3 modification. These regions are found to be highly enriched with ATRX and H3.3. Remarkably, loss of ATRX leads to a reduction of the histone variant and H3K9me3 levels, resulting in activation of the associated silenced imprinted genes. This suggests that ATRX deposition of H3.3, presumably facilitated by DAXX, at DMRs is essential for repression of imprinted genes (Voon *et al*., 2014 submitted).

1.3.2.3 ATRX-mediated chromatin association of macroH2A1

MacroH2A histone variant belongs to H2A histone protein family and is characterised by the presence of a C-terminal macro domain (Pehrson and Fried 1992). There are two isoforms, encoded by two genes, macroH2A1, which has two splicing isoforms macroH2A1.1 and macroH2A1.2, and macroH2A2 (Costanzi and Pehrson 2001). This histone variant is generally transcriptionally repressive as it has been shown to hinder accessibility of transcription factors to DNA and produce inflexible nucleosomes. It has also been found in senescence-associated heterochromatic foci and inactivated X chromosome (Costanzi and Pehrson 1998, Zhang, Poustovoitov et al. 2005).

Bernstein's group investigated how macroH2A is deposited into chromatin by ATRX. Purification of chromatin-free GFP-tagged H2A and GFP-tagged macroH2A, followed by immunobloting showed that ATRX is associated with this histone variant in a distinct complex rather than that of H3.3. Furthermore, stable knock-down of ATRX using shRNAs increased the level of macroH2A1 in chromatin, particularly at telomeres and α-globin locus, without changing global expression of the histone

variant. This suggests that ATRX negatively regulates the incorporation of macroH2A1 into chromatin. However, it is unclear whether ATRX directly associates with macroH2A1 to evict this histone variant from nucleosomes or prevents its deposition to chromatin (Ratnakumar, Duarte et al. 2012).

The general presence of macroH2A in heterochromatin suggests that it might be recruited to telomeres by ATRX to maintain genome integrity. The increase of macroH2A localisation at the α-globin locus in the absence of ATRX might provide a compelling explanation for the down-regulation of α-globin expression in ATRX syndrome although further study into a mechanism is necessary (Ratnakumar, Duarte et al. 2012). Furthermore, additional factors might be needed to explain for the variation in reduction of the gene expression.

In addition to the regions that have been discussed above, what are other locations in the genome and the nucleus, in general, ATRX is localized, and how does it recognise a specific target location?

1.3.2.4 Localisation of ATRX and its targeting signal

1.3.2.4.1 PML nuclear bodies

PML-NBs (Promyelocytic leukemia nuclear bodies) are speckles that are present in the nuclear matrix. PML protein, which is involved in chromosomal translocations in acute PML, is the key organiser of these super structures. PML-NBs are known to recruit numerous nuclear proteins and are believed to play roles in many nuclear processes including transcriptional activation, DNA replication, apoptosis and viral infection (Lallemand-Breitenbach and de The 2010).

Immunofluorescene staining shows that ATRX colocalises with DAXX and H3.3 in PML-NBs (Xue, Gibbons et al. 2003). PML-NBs serves as a platform to tether ATRX/DAXX/H3.3 to telomere in mouse ES cells, a process important for maintaining telomere integrity (Chang, McGhie et al. 2013). Furthermore, it was found that the C-terminal domain of ATRX is responsible for the localisation and that ATRX mutations in this domain disrupt the association with PML-NBs (Berube, Healy et al. 2008).

1.3.2.4.2 Pericentric heterochromatin localisation of ATRX is dependent on H3K9me3

Pericentric heterochromatin is the region comprised of A/T rich "major" satellites, which flanks the centric heterochromatin consisting of "minor" satellite repeats. Pericentric heterochromatin is characterised by methylated CpG sites, a low level of acetylated histone modifications, and especially a high level of H3K9me3 and HP1 (Heterochromatin Protein 1) (Probst and Almouzni 2008).

As discussed in the previous section, ATRX localises at pericentric heterochromatin to deposit H3.3 histone variants, the process assisted by DAXX histone chaperone. What specific features of this region does ATRX recognise for its localisation? The initial finding is that ATRX targeting to pericentric heterochromatin involves an interaction with HP1 (Le Douarin, Nielsen et al. 1996). Furthermore, evidence showed that this targeting requires H3K9me3 as when Suv39H1 and Suv39H2 methyltransferases, enzymes that are responsible for writing this histone mark, are deleted, ATRX localisation is completely abolished (Eustermann, Yang et al. 2011).

An appealing mechanism proposed from these two observations is that ATRX interacts with the C-terminal chromo shadow domains of HP1α and HP1β, which recognise K9me3 through their N-terminal chromodomains (Nielsen, Nietlispach et al. 2002). However, one argument against this hypothesis is that it provides little specificity for the targeting since HP1 interacts with many other proteins (Lechner, Schultz et al. 2005).

This problem is solved by two independent studies by Eustermann *et al*. and Iwase *et al*. They simultaneously published their elegant work proving that H3K9me3 serves as a direct targeting signal for the localisation of ATRX to pericentric heterochromatin. More particularly, the ADD domain of ATRX forms two binding pockets, one specifically accommodating unmethylated Lys4 (H3K4) and the other one restricted to di- and tri-methylated Lys9 (H3K9me2 and H3K9me3). This combinatorial readout is required for ATRX recruitment, which is further enhanced by the interaction through HP1 that also recognises H3K9me3 (Eustermann, Yang et al. 2011, Iwase, Xiang et al. 2011).

1.3.2.4.3 Telomeres and ribosomal DNA repeats

A telomere is a region of repetitive sequence at each end of a linear chromosome, playing essential roles in chromosome replication and genome stability. In vertebrates, telomeres are several kb arrays of G-rich short tandem repeats that has the sequence TTAGGG (see review (O'Sullivan and Karlseder 2010)).

Ribosomal DNA repeats (rDNA) are also GC-rich repeats. Mutations in ATRX cause these repeats to become hypomethylated (Gibbons, McDowell et al. 2000).

As discussed in the previous section, ATRX, together with DAXX, is recruited to telomeres to deposit H3.3. In addition, it was also reported that ATRX binds rDNA with maximal binding occurring at the transcribed region of the locus that is very G-rich and CpG nucleotides (McDowell, Gibbons et al. 1999) (Law, Lower et al. 2010). In our laboratory's most recent ATRX-ChIP seq in wild-type and a double Suv39H1/H2 knock-out of mouse embryonic fibroblasts (MEFs) showed that loss of H3K9me3 does not decrease, but rather slightly increase (by 1.27 fold), the level of ATRX binding at telomeres (Hsiao Voon, personal communication). Likewise, the levels of ATRX enrichment at rDNA in the wild-type and double knock-out cells are similar (Hsiao Voon, personal communication). This suggests that unlike A/T rich pericentric heterochromatin, ATRX targeting to GC-rich telomeres and rDNA repeats is independent of H3K9me3, suggesting that different features in these regions might be required for ATRX localisation.

A recent study about the roles of DEK protein in regulating H3.3 deposition into chromatin may have provided a clue to what signals ATRX targeting to telomeres. In this study, the oncoprotein DEK, another histone chaperone for H3.3 was shown to associate with PML-NBs, together with ATRX at telomeres in embryonic stem (ES) cells. Depletion of DEK in ES cells results in disassociation of PML and ATRX with telomeres, diminishing H3.3 at these sites and broadly redistributing H3.3 to chromosome arms and pericentric heterochromatin. As a consequence, telomeres in DEK-deficient cells become fragile, aberrant in structures and enriched in DNA damage mark γH2A.X (Ivanauskiene, Delbarre et al. 2014). These data suggest that DEK is required for targeting of ATRX, and deposition of H3.3 at telomeres.

Interestingly, DEK has been shown to bind DNA with no sequence specificity but with a preference for non-canonical DNA conformations such as supercoiled or cruciform DNA (Waldmann, Scholten et al. 2004, Bohm, Kappes et al. 2005). Telomeres are G-rich repeat regions that have been shown to form secondary structures, such as G-quadruplexes and R-loops. It is tempting to speculate that DEK directly binds to these structures at telomeres, recruiting ATRX for H3.3 deposition at these regions.

1.3.2.4.4 CpG islands and GC-rich tandem repeats

Genome-wide analyses of ATRX-ChIP seq data show that ATRX also targets CpG islands and GC-rich tandem repeats (Law, Lower et al. 2010) (Voon *et al*., 2014 submitted). Interestingly, ATRX recruitment to these regions are unaffected by a double Suv39H1/H2 knock-out. Remarkably, K9me3 requirement for ATRX localisation to GC-rich DNA repeat sequence seems to be inversely proportional to the GC content (Hsiao Voon, 2014, personal communication). These observations suggest that H3K9me3 might not be targeting signal for ATRX recruitment to CpG islands and GC-rich tandem repeats.

An obvious question to ask at this point is if the well-characterised H3K9me3 interaction is not required for ATRX localisation to the GC-rich targets, including telomeres, rDNA repeats, GC-rich tandem repeats and CpG islands, what is? GC-rich repeats are prone to form non-canonical DNA structures, which will be discussed in the following section. An appealing hypothesis to emerge from these observations is that ATRX may use these structures as a targeting signal. In support of this, ATRX has been shown to bind G-quadruplexes, a four-stranded DNA structure, *in vitro*

(Law, Lower et al. 2010). Whether they are indeed targeting signals for ATRX recruitment remains to be elucidated.

1.3.3 The role of ATRX in gene regulation: ATR-X syndrome and ATMDS

Inherited mutations of ATRX result in a rare developmental disorder, called ATR-X syndrome. Most mutations found in patients fall into the two known domains: 50% of the mutations are located in the ADD domain and 30% occur in the ATPase domain(Gibbons, Wada et al. 2008). The patients display a range of phenotypic abnormalities including facial dysmorphism, genital abnormalities, learning difficulties and alpha thalassemia (a specific form of anemia caused by reduction in alpha globin production) (Gibbons 2006). This suggests that ATRX regulates expression of many genes, including the alpha globin genes. Further supporting the idea that ATRX plays a direct role in regulating the expression of the alpha globin genes, patients with ATMDS (Alpha Thalassemia associated Myelodysplasia Syndrome) have been shown to possess acquired ATRX mutations and profound down-regulation of alpha globin expression (Gibbons, Pellagatti et al. 2003).

Although mutations of ATRX are linked to decreased alpha globin expression, the severity of the dysregulation is independent of the specific mutations in ATRX. Comparison of 32 cases from 26 pedigrees with the same common mutation in ATRX (736C>T) shows that the level of alpha thalassemia, measured by the percentage of red cells containing HbH inclusions (formed from β4), varies in these individuals, ranging from 0-14% (Gibbons 2006). The fact that the patients with identical mutations might display very different, albeit stable, degrees of alpha globin

expression level indicates that there might be other genetic factors that contribute to the regulation of alpha globin expression (Gibbons 2006).

The exact mechanism by which ATRX controls alpha globin expression is still unclear although a recent finding by Law and colleagues may help to address this question. Previous studies have demonstrated enrichment of ATRX at ribosomal DNA repeats, Y-specific satellite, heterochromatic repeats and sub-telomeric repeats (McDowell, Gibbons et al. 1999, Gibbons, McDowell et al. 2000, Xue, Gibbons et al. 2003), but Law *et al*. were able to determine the genome-wide binding of ATRX in human and mouse cells. This study showed that ATRX binds to G-rich tandem repeats in or near the genes whose expression is perturbed in ATR-X patients. Interestingly, in the terminal region of the 16p chromosome, there are three peaks of ATRX and all have a high G content. A small and reproducible peak was seen in the proximity of the telomere of 16p. The second peak was found at 1kb upstream of HBM promoter, at ψζVNTR (GCGGG)n. The third peak is located at a VNTR in the first intron of NME4 gene (Law, Lower et al. 2010). What is noticeable here is that the peaks at the NME4 VNTR and 16p telomere are consistently found in different cell types, including erythroid and non-erythroid cell types, whereas the peak at ψζVNTR is only seen in erythroid cells. Given that ψζVNTR is located in ψζ gene, which has the potential to be transcriptionally active in erythroid cells but not active in non-erythroid cells, one implication of this could be that the transcription process plays a role in ATRX recruitment to its targets.

The ψζVNTR is highly polymorphic. Interestingly, the study by Law and coworkers showed a significant correlation between the size of the tandem repeat in patients and

the severity of alpha thalassemia (reflecting the degree of alpha globin reduction) measured by the percentage of cells with HbH inclusions. The larger the tandem repeat, the greater the reduction of alpha globin expression. The degree of down-regulation of affected genes was found to be related to their proximity to the peak of ATRX binding. How ATRX regulates gene expression and how the size of the VNTR influences this process remains to be determined.

1.3.4 The roles of ATRX in DNA replication and genome instability

As previously discussed, ATRX binds interstitial tandem repeats and telomeres (Law, Lower et al. 2010) during S-phase and in the absence of ATRX there is an increase DNA damage at telomeres (Wong, McGhie et al. 2010). The important findings raise the possibility that ATRX may be required for DNA replication at these regions and loss of ATRX leads to replicative stress.

Indeed, ATRX deficient mouse ES cells show defects in DNA replication including an increase in sensitivity to DNA damage agent hydroxyurea, a prolonged S-phase, and an increase in stalled replication fork and double-strand breaks (Huh, Price O'Dea et al. 2012) (Leung, Ghosal et al. 2013, Clynes, Jelinska et al. 2014). Furthermore, phosphorylated CHK1 (checkpoint 1) is greatly reduced in ATRX-null mouse ES cells treated with a low dose of hydroxyurea, resulting in slower progression to mitotic entry. This suggests that ATRX is required for checkpoint activation in response to DNA damage. In addition, ATRX is shown to co-localise with single-stranded DNA binding protein RPA2 at DNA damage sites that are caused by laser-induced micro irradiation, suggesting that ATRX directly participates in replication stress pathway (Leung, Ghosal et al. 2013).

Given that ATRX plays an important role in DNA replication, one might expect that loss of ATRX will have consequences in cellular processes such as genome instability, chromosomal segregation defects and compromised growth. Indeed, depletion of ATRX is associated with extensive genomic rearrangements and telomere fusions (Berube, Mangelsdorf et al. 2005) (Huh, Price O'Dea et al. 2012). Furthermore, loss of ATRX functions has been reported to cause defective spindle formation and chromosome condensation and alignment in mitosis and meiosis (Ritchie, Seah et al. 2008) (De La Fuente, Viveiros et al. 2004). In addition, ATRX-deficiency in myoblast, Sertoli, and neuprogenitor cells is associated with compromised muscular growth, testicular defects and increased neuronal loss, respectively (Bagheri-Fam, Argentaro et al. 2011, Huh, Price O'Dea et al. 2012) (Berube, Mangelsdorf et al. 2005).

Further studies aiming to understand the mechanism by which ATRX prevents replication stress and genome instability show that ATRX directly associates with components of the MRN complex (Leung, Ghosal et al. 2013) (Clynes, Jelinska et al. 2014). More specifically, it was shown that the co-localisation of ATRX and MRN occurs during S phase (Clynes, Jelinska et al. 2014). The MRN complex functions in DNA damage repair (both via homologous recombination and non-homologous end joining) and the restart of stalled replication forks (Robison, Elliott et al. 2004) (Bryant, Petermann et al. 2009). Taken together, these observations provide evidence that ATRX interacts with MRN complex, and may directly facilitate the DNA replication process (Clynes, Jelinska et al. 2014).

1.3.5 ATRX and cancer

The first clue to the link between ATRX and cancer comes from the observation that a cohort of patients with myelodysplastic syndrome (MDS) concurrently acquired somatic mutations in ATRX associated with alpha-thalassemia (ATMDS) (Steensma, Higgs et al. 2004) (Gibbons, Pellagatti et al. 2003). It was, however, unclear as to whether ATRX is a passenger mutation or a driver event in ATMDS.

In order to survive and multiply, cancer cells have to maintain telomeres. In most of cancers, this task relies on overexpression of telomerase, the enzyme that adds telomeric sequence to chromosome ends every round of DNA replication (Shay and Bacchetti 1997, Blackburn, Greider et al. 2006). However, in 10-15% cancers which lacks detectable telomerase, an alternative mechanism is used, called alternative lengthening telomere (ALT) pathway, which relies on homologous recombination between telomeric sequences (Cesare and Reddel 2010, Killela, Reitman et al. 2013) (Shay, Reddel et al. 2012). Strikingly, a recent study has shown that ATRX is lost in 90% of *in vitro* immortalised cell lines that use ALT pathway for survival (Lovejoy, Li et al. 2012). Moreover, mutations in ATRX and/or DAXX have been found in many cancers displaying ALT, including pancreatic neuroendocrine tumors, neuroblastomas, paediatric glioblastomas, oligodendrogliomas, and medulloblastomas (Heaphy, de Wilde et al. 2011, Lovejoy, Li et al. 2012, Schwartzentruber, Korshunov et al. 2012). These observations are compelling evidence that ATRX is a suppressor of the ALT pathway. Interestingly, knock-out of functional ATRX alone is insufficient to trigger ALT in mouse ES cells and HeLa cells, suggesting that additional factors are needed to activate this pathway (Clynes, Jelinska et al. 2014) (Lovejoy, Li et al. 2012).

1.4 NON-B FORM DNA STRUCTURES IN THE GC-RICH REGIONS OF THE GENOME

The right-handed double helix structure of DNA was discovered in 1953. Most of DNA in the cells are in this canonical B form. However, it has become clear that DNA can adopt a wide range of alternative non-canonical conformations based on sequence motifs and interactions with proteins. Examples are Z-DNA, cruciform, triplex, hairpin and G-quadruplex. GC-rich sequences have been demonstrated to form G-quadruplex conformation. Furthermore, these regions are also prone to adopt R-loop structures when interacting with nascent RNA transcripts during transcription.

1.4.1 G-quadruplexes

1.4.1.1 Structure of G-quadruplexes (G4)

One of the very first G-quadruplex structures was described by Gilbert's laboratory when examining the immunoglobulin switch region in 1988. They discovered that single-stranded DNA containing short G-rich tandem repeats will self-associate to make four-stranded structures which are held together by guanines bonded to each other by Hoogsteen pairing (Sen and Gilbert 1988). The sequence motif that predicts G4 formation contains at least four runs (or clusters) of at least three guanine bases, and these G-runs are usually separated by other bases although the formation is also possible without this spacer base: $G_{+3}N_xG_3N_xG_3N_xG_3$. One G4 is a stack of three or more G-quartets (basic units of G4), a square planar assembly of four guanine bases, and the intervening sequences, which are the spacer bases (N), are extruded as single-strand loops (Bochman, Paeschke et al. 2012) (Tarsounas and Tijsterman 2013). G4

structures with smaller loops and/or longer G-runs are more stable (Duquette, Handa et al. 2004) (Huppert 2010).

G4 can adopt a variety of topologies depending on the directionality of the G-runs making up the core structure. These may be parallel, antiparallel or hybrid. In addition, a G4 complex may form using guanines from the same nucleic acid strand (intramolecular) or from different strands (intermolecular). Furthermore, the connecting loops of G4 can vary in length and sequence which influence the topology of the structure.

There is also evidence supporting the presence of RNA G4s. They are even more readily formed since RNA is single-stranded and so G4 formation in this case does not have to compete with duplex formation like it does with DNA strands. RNA G4 structures are more stable than the structures formed by DNA (Hofacker 2003) (Huppert 2010).

G4 structures are stabilized by a K+ or Na+ ion in the centre of the quartet at physiological temperature and pH in vitro (Wu and Brosh 2010). These complexes are much more stable than DNA duplexes, with 20-30^{o}C higher melting temperature, suggesting that specific proteins are required to unfold these highly stable structures (Lipps and Rhodes 2009).

Intramolecular G4 quadruplexes are more likely to form in the circumstances when DNA is transiently denatured, such as during DNA replication, transcription and repair (Lipps and Rhodes 2009). One idea is that these structures may impede normal

functioning of these cellular processes and, hence, need to be resolved in order to complete the process in question. This raises the question of whether these processes are prone to dysregulation in the absence of putative proteins that resolve these structures.

1.4.1.2 In vivo existence of G4

Although G4 DNA has been proven to exist in cell-free systems, the existence of G-quadruplexes in living cells has been controversial until recently. One of the first elegant pieces of evidence for *in vivo* G4 structures comes from a study in which high affinity antibodies raised against G4s bind specifically to the telomeres of the ciliate *Stylonichia lemae* (Schaffitzel, Berger et al. 2001, Qin and Hurley 2008). Recently, several groups have developed antibodies specific for G4 structures in mammalian cells. The murine monoclonal antibody 1H6 has been used to visualise G4s in HeLa and DT40 cell lines using immunofluorescence staining technique (Henderson, Wu et al. 2014). Remarkably, immunofluorescence imaging with the single-chain antibody BG4, developed by the Balasubramanian laboratory, shows DNA G4 structures at telomeres, at interstitial sites across chromosomes (Biffi, Tannahill et al. 2013) and RNA G4 structures in the cytoplasm of human cells (Biffi, Di Antonio et al. 2014). Moreover, another antibody, called hf2, has recently successfully been used to enrich G-quadruplexes for deep-sequencing and mapping G4 structures at high resolution in genomic DNA of human breast adenocarcinoma cells (Lam, Beraldi et al. 2013).

A large number of G4 ligands have been developed to help address the question of the existence of G-quadruplexes in living organisms. These small molecules are able to

recognise and stabilise G4 structures *in vitro*. Examples are pyridostatin, telomestatin, PhenDC3 and TMPyP4. Cells treated with these G4 ligands show detrimental phenotypes such as DNA damage at telomeres, recombination-dependent rearrangement at minisatellites, genomic instability and growth inhibition (pyridostatin, telomestatin, PhenDC3) (Rodriguez, Muller et al. 2008) (Muller, Sanders et al. 2012) (Gomez, Wenner et al. 2006) (Piazza, Boule et al. 2010) or down-regulation of the expression of genes containing G4 motifs in the promoter (TMPyP4) (Grand, Han et al. 2002), indirectly suggesting the cellular existence of G4 structures.

Publications about the biological relevance and functions of G4 DNA in living cells have increased dramatically in the past decades. These have been accompanied by rising numbers of proteins that have been found to target G4 DNA. For example, a growing group of helicases, including Pif1, FANCJ, BLM, and WRN have been demonstrated to bind specifically to G4 DNA and resolve the structures *in vitro* (Sun, Karow et al. 1998, Fry and Loeb 1999, Hickson, Davies et al. 2001, Mohaghegh and Hickson 2001, Mohaghegh, Karow et al. 2001, Paeschke, Capra et al. 2011, Sarkies, Murat et al. 2011). Mutations of these cause genomic instability and human disease, supporting the *in vivo* relevance of G-quadruplexes.

1.4.1.3 Genomic location of G-quadruplexes

Sequence motifs have been found throughout the genome. Computational analyses of the entire human genome reveal >375,000 G4 motifs whereas there are >1400 G4 sequences in the nuclear genome of *S. cerevisiae*. These appear to occur non-randomly (Lipps and Rhodes 2009), and tend to concentrate at functional genomic

domains, such as telomeres, rDNA arrays, immunoglobulin heavy chain switch regions and minisatellites.

Moreover, G quadruplex elements have been found in the promoter regions of a number human genes, such as c-MYC (Brooks and Hurley 2010), VEGF (Sun, Liu et al. 2008), HIF-1α (Guan, Reddy et al. 2010) and KRAS (Cogoi and Xodo 2006), and Bcl-2 (Dai, Dexheimer et al. 2006). The advantage of localising in DNase I Hypersensitive Sites (DHS) near Transcription Start Site (TSS) of most of these G4 motifs is that these sites are nucleosome-free and negative supercoiling torque is the highest, which causes the DNA duplex to melt, promoting G4 formation (Balasubramanian, Hurley et al. 2011). These G4 motifs have been demonstrated to play a critical role in transcriptional regulation.

A variety of different G4 prediction tools have been developed (Huppert and Balasubramanian 2005) (Todd, Johnston et al. 2005) (Eddy and Maizels 2006) (Todd 2007) (Huppert 2008). One of which, the QUADPARSER, is widely used in G4 studies and identifies the sequence of the form $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ (Huppert and Balasubramanian 2005).

1.4.1.4 Studying G4 structure in vitro

A number of different techniques have been employed to study the formation and structure of G-quadruplexes. They allow us to examine different aspects of the structure. The first technique was used to determine the complete structure of G4 is X-ray crystallography (Gellert, Lipsett et al. 1962). However, significant

shortcomings of X-ray crystal structures are that (1) only the G4s formed in solid state can be investigated, which may be different from the structure formed in solution, and (2) only the most easily adopted G4 configuration is described whereas in fact G4 structures are highly polymorphic (Huppert 2008). NMR spectroscopy, another technique for detailed structure determination can improve these, allowing one to study dynamics, kinetics and polymorphism of G-quadruplexes (Webba da Silva 2007).

Circular dichroism (CD) spectroscopy is the simplest way of predicting G4 presence (Huppert 2008). It also allows discrimination of all-parallel and anti-parallel topologies of G-quadruplexes (Law, Lower et al. 2010) (Gonzalez, Guo et al. 2009). In addition, fluorescence spectroscopy and FRET (fluorescence resonance energy transfer) have been used to study the folding of G4 structure (Huppert 2008) (Paramasivam, Membrino et al. 2009).

DMS (Dimethylsufate) footprinting is a useful technique to determine the presence of G4 (Duquette, Handa et al. 2004) (Gonzalez, Guo et al. 2009). This technique is based on the fact that DMS methylates N_7, which leads to cleavage at this position in the DNA sequence. However, because in a G-quadruplex the N_7 of guanine participates in a hydrogen bond it is protected from methylation by DMS, resulting in no cleavage (Huppert 2008).

In addition, the stability of G-quadruplexes reduces when they go from a higher temperature to a lower temperature, causing a melting transition between the two

temperatures. This characteristics can be monitored by the absorbance of UV light at 295nm (Mergny, Phan et al. 1998).

1.4.1.5 Helicases that resolve G4 structure

Cells have evolved a large group of helicases that unwind G-quadruplexes. Bochman *et al*. reported that 22 different helicases have been tested for their ability to bind and/or resolve G4 structures in various studies. All but one, the *E. coli* RecBCD helicase, were positive, suggesting that G4 unwinding is a non-specific activity for many helicases (Bochman, Paeschke et al. 2012). It is unclear, however, if different G4 helicases target different G4 configurations.

A large number of G4 unwinding helicases belong to the RecQ family, a highly-conserved enzyme family homologous to *E.coli* RecQ. These enzymes unwind G4s in 3'->5' direction, and include RecQ in *E. coli*, Sgs1 in *S. cerevisiae*, Rqh1 in *S. pombe*, WRN and BLM in chicken and WRN, BLM, RTS (RecQ4) in human. Among these, WRN and BLM helicases are the most extensively studied. Apart from G4 structures, they also unwind a wide range of substrates including 3'-tailed duplexes, bubble structures, forked duplexes, DNA displacement loops and four-way (Holliday) junctions. Mutations in WRN, BLM and RecQ4 have been found in human genetic disorders Werner's syndrome (WRN), Bloom's syndrome (BLM) and RTS, RAPADILINO syndrome and BGS (RecQ4) (review (Chu and Hickson 2009)).

In addition, the DOG-1 family of 5'->3' DNA helicases has been reported to resolve G4 structures in many studies. The FANCJ (also called BACH1 or BRIP1) member of

this family was previously shown to efficiently unwind duplex and forked DNA by Gupta *et al*. (Gupta, Sharma et al. 2005); however in a re-investigation by a different group it has been shown to preferentially unwind a G4 DNA substrate, but not duplex and forked DNA *in vitro* (London, Barber et al. 2008) (Wu, Shin-ya et al. 2008). In support of the specificity of FANCJ for G4 structure, FANCJ is required for replication progress *in vivo* in the presence of the G4 stabiliser telomestatin (Schwab, Nieminuszczy et al. 2013) or PhenDC3 (Castillo Bosch, Segura-Bayona et al. 2014). Deletion of FANCJ causes replication fork stalling at G4 sites (Castillo Bosch, Segura-Bayona et al. 2014), impaired proliferation and DNA damage (Wu, Shin-ya et al. 2008). Other members of the DOG-1 family that have been demonstrated to target and resolve G4s include nematode DOG-1, mouse RTEL, human XPD and ChlR1 (Jones and Rose 2012, Wu, Sommers et al. 2012, Gray, Vallur et al. 2014) (Vannier, Sandhu et al. 2013).

Yeast and human Pif1, a distant relative of 5'->3' RecD bacterial helicase, have also been shown to have G4 unwinding activity *in vitro* and *in vivo* (Ribeyre, Lopes et al. 2009) (Sanders 2010), preventing genome instability caused by G-quadruplex formation. Pif1 helicase has also been demonstrated to remove R-loops (Boule and Zakian 2007) (Paeschke, Capra et al. 2011) (Zhou, Zhang et al. 2014).

1.4.1.6 Other G4-interacting proteins

In addition to the G4-unwinding helicases, other proteins have been shown to interact with G4 structures to either unfold or stabilise them. Nucleolin, an essential nucleolar protein involved in homologous recombination repair, has been shown to interact

directly with G-quadruplexes *in vitro* and *in vivo* (Gonzalez, Guo et al. 2009) (Indig, Rybanska et al. 2012) (Haeusler, Donnelly et al. 2014). It competes with WRN binding of G4 tetraplex DNA, inhibiting the G4 unwinding helicase activity of WRN (Indig, Rybanska et al. 2012). In another study, nucleolin is demonstrated to facilitates the formation and increase the stability of the G-quadruplexes in the c-MYC promoter (Gonzalez, Guo et al. 2009).

Evidence from a pull-down experiment using a nuclear extract and beads coupled with G4s suggests that the nuclear proteins Ku70, hnRNP A1 and PARP-1 are associated with G-quadruplexes in the G-rich promoter of the KRAS oncogene (Cogoi, Paramasivam et al. 2008). Further experiments demonstrated that hnRNP A1 and its derivative UP1 unfold quadruplex DNA in KRAS and minisatellite repeats (Zhang, Manche et al. 2006, Paramasivam, Membrino et al. 2009) (Fukuda, Katahira et al. 2002).

MutSα (Msh2/Msh6), a mismatch repair complex, is another example of a G4 binding protein. It has high affinity with the G-quadruplexes formed in the transcribed S regions in B cells, promoting DNA recombination (Larson, Duquette et al. 2005). MutSα was also shown to inhibit the G4 unwinding activity of FANCJ helicase (Wu, Shin-ya et al. 2008).

In contrast to MutSα, the human shelterin protein POT1 disrupts G-quadruplexes formed at telomeres, facilitating telomerase extension *in vitro*. It was demonstrated that the DNA-binding domain at the N terminal region of POT1 is sufficient for this activity (Zaug, Podell et al. 2005).

The chromatin remodeling protein ATRX has been shown to be enriched at G-rich tandem repeats in the mouse and human genome. Furthermore, it was reported to bind G4 quadruplexes *in vitro* in gel shift assays (Law, Lower et al. 2010). However, it is unclear as to whether ATRX binds this structure *in vivo* and which part of the protein is responsible for the binding. In addition, unlike the G4-binding proteins discussed above, why ATRX interacts with G-quadruplexes is unknown. It has recently been shown to be unable to unwind G4s *in vitro* (Clynes, Jelinska et al. 2014).

1.4.1.7 Biological functions of G4

1.4.1.7.1 G-quadruplexes physically block transcription and replication processes

Intramolecular G-quadruplexes are more likely to form in the circumstances when DNA duplex is transiently melted into two single strands, such as during DNA replication, transcription and repair (Lipps and Rodes. 2009). If not resolved, they can impede these cellular processes.

G4 forming sequences have been shown to arrest transcription by both T7 polymerase and mammalian RNA polymerase II *in vitro*, and this is dependent on the ability to form G-quadruplexes (Tornaletti, Park-Snyder et al. 2008) (Broxson, Beckett et al. 2011). However, it is controversial whether the arrest occurs only when G4 sequences are on the non-transcribed or the transcribed strand. According to Tornaletti, the inhibition of trancription occurs only when the G tracts are located on the non-transcribed strand, but not on the transcribed-strand (Tornaletti, Park-Snyder et al. 2008), whereas the study on the G4 motif of the human *c-myb* gene by Broxson resulted in the opposite conclusion with a suggestion that the G-tetrads act as a

roadblock to the RNA polymerase (T7) movement, impeding the elongation process when it reaches the catalytic site (Broxson, Beckett et al. 2011). Perhaps, the mechanisms of transcription arrest are different in the two cases. The G4 forming sequence inserted in the plasmid in the first study was shown to form a G-loop which includes G4s on the non-transcribed strand and an RNA:DNA hybrid on the other strand, which might be the direct cause of the arrest.

There is better evidence showing DNA replication blockage by G4 structures *in vivo*. In most cases, the blockage happens in the absence of G4-unwinding helicases. Using DNA polymerase II high occupancy to monitor replication fork pausing, it is shown that pif1 mutants are associated with pausing near the sites of G4 forming sequences (Paeschke, Capra et al. 2011). Folding of G-quadruplexes is also observed to hinder replication fork progress at ribosomal DNA genes in *S. cerevisae*, which are enriched with G4 motifs and PIF1 is required for the DNA polymerase to overcome these sites (Brewer, Lockshon et al. 1992, Azvolinsky, Giresi et al. 2009). In addition, the replicating plasmid assay shows that G4 motifs on the leading strand lead to a marked reduction in the efficiency of replication. Moreover, REV1 polymerase overexpression restores the reduced efficiency to the normal level (Sarkies, Reams et al. 2010).

1.4.1.7.2 Gene regulation through G4 motifs

A hint for the role of G-quadruplexes in gene regulation lies in the fact that G4 motifs are enriched in promoter regions near transcription start sites (Huppert and Balasubramanian 2007). More than 40% of human gene promoters contain one or more G4 motifs that strongly associate with DNA hypersensitive sites (DHS) and are

structurally biased towards more stable forms, suggesting direct roles for these motifs in gene regulation (Huppert and Balasubramanian 2007). Examples include a large number of protooncogenes, such as c-MYC, VEGF, HIF-1α, KRAS, and Bcl-2, as well as other genes reviewed in (Qin and Hurley 2008).

The G4 motif in NHEIII of the c-MYC promoter is shown to form G-quadruplexes *in vitro*. More importantly, these structures serve as a regulatory element to which effector proteins such as nucleolin and NM23-H2 bind, and through stabilising and unfolding it, respectively, the expression of the gene is modified (Brooks and Hurley 2010). Another example is the human KRAS promoter that contains a sequence that can form two different G-quadruplex structures (Cogoi, Paramasivam et al. 2008). In this case, the effector proteins are Ku70, PARP1 and hRNP A1 that have been discussed in the previous section.

Since G4 motifs are demonstrated to play important roles in the regulation of many proto-oncogenes, they are promising targets for drugs in cancer treatment.

G4 sequences are also found enriched in the first intron of human genes, where they are thought to provide structural targets for regulation at the transcriptional level of RNA processing (Eddy and Maizels 2008).

1.4.1.7.3 Epigenetic instability induced by G-quadruplexes

To maintain epigenetic memory through cell division, the cells rely on a tight coupling between DNA synthesis and histone recycling, in which the redeposition of parental histones will be used to copy faithfully epigenetic marks onto the newly

incorporated histones. G-quadruplexes have been shown to slow and block replication machinery *in vitro* (Woodford, Howell et al. 1994) and *in vivo* (Schaffitzel, Berger et al. 2001, Sfeir, Kosiyatrakul et al. 2009), which is normally overcome by a process called translesion DNA synthesis (TLS) to help maintain the faithful deposition of epigenetic marks on newly synthesised DNA (Lopes, Foiani et al. 2006, Sarkies, Reams et al. 2010). But what happens if this TLS pathway is impaired?

The Sale group observed that when players of the TLS pathway, such as REV1-a member of the Y family of DNA polymerase, are mutated, G-quadruplexes formed at genes are bypassed by the replication machinery, leaving a gap around the G4 sequence, which will be filled by DNA synthesis behind the fork. As a consequence of this, histone recycling is uncoupled from bulk DNA synthesis, resulting in loss of parental chromatin marks and predominant incorporation of new histones. This causes loss of epigenetic memory, which explains the derepression of transcriptionally inactive and silencing of active loci containing G4 motifs (Sarkies, Reams et al. 2010) (Sarkies, Murat et al. 2011). This phenomenon is described in detail in the case of the chicken β-globin gene, which contains three G4 sequences in the second intron and is normally transcriptionally inactive, becomes active in the rev1-deficient cells due to loss of parental H3K9me2 and an increase in marks of new histone deposition.

Further work by the Sale group demonstrated that G4-induced epigenetic instability is also caused by loss of function of G4 unwinding helicases such as FANCJ, BLM and WRN (Sarkies, Murat et al. 2011).

Interestingly, this disruption of the accurate propagation of chromatin structure is dependent on the G4 motifs residing on the leading template strand, not the lagging strand. Moreover, the loss of chromatin modifications around TSS correlates with the position of the G4 motifs (Schiavone, Guilbaud et al. 2014).

1.4.1.7.4 Translation regulation by G4 RNA

Recently, G4 motifs have been found to mark elF4A-dependent mRNAs. ElF4A is a RNA helicase that is up-regulated in cancer. It is able to transform fibroblasts and promote tumour development *in vivo* by initiating cap-dependent translation of oncogene mRNAs, such as RAS and ERK. Therefore, elF4A has been used a target for cancer drugs such as silverstrol and related compounds. Using ribosome footprinting technology, Wolfe *et al*. was able to perform deep sequencing all the target genes that are sensitive to the elF4 inhibitor, silverstrol. Strikingly, they found that a significant number of affected genes are enriched for $(CGG)_4$ and $(CGG)_3$ motifs at their 5'UTR and these can fold into G-quadruplex RNA structure as detected by *in vitro* assays. Similarly to the silverstrol effect, elF4A knock-down results in a decrease in the protein level of a reporter harbouring G4 motifs in the 5' UTR. These observations suggest a new role of G4 RNA in a translational controlling pathway dependent on elF4A (Wolfe, Singh et al. 2014).

1.4.1.7.5 G4 sequences serve as replication origins

DNA replication is initiated at specific sites in the genome, called replication origins. Previously, little was known about the position and characteristics of the replication origins in the human genome. In an attempt to close this gap, Besnard and colleagues

extensively mapped the genome-wide location of the origins in 4 different cells types using deep-sequencing of short nascent strands. They identified ten times more replication origins than expected and most of them are conserved among the cell types. Notably, a subset of the origins are associated with G4 motifs with their density being correlated with the origin density and timing of replication, suggesting that G4 sequences might play a role in origin regulation (Besnard, Babled et al. 2012).

A recent study has provided an insight into the mechanism of the G4 origin regulation. Using model origins, it is shown that G4 motifs are required for replication initiation. In one case, two G4 structures cooperate to induce replication initiation and in the other case, an additional *cis*-regulatory element was required to start replication. Furthermore, the potential of G4 formation was found to correlate with the efficiency of initiation (Valton, Hassan-Zadeh et al. 2014).

1.4.1.7.6 G4 mediated genome instability

G-quadruplexes can obstruct DNA replication fork movement, which can lead to DNA breakage, triggering DNA recombination, rearrangement and genome instability. G4-induced genome instability typically occurs when the cells fail to unwind G4 structures, leading to the road blocks to the replication machinery. For example, G4 DNA sequences that match the signature $G_{3-5}N_{1-3}G_{3-5}N_{1-3}G_{3-5}N_{1-3}G_{3-5}$ are deleted from the genome of *C. elegans* cells that lack DOG-1, a G4 unwinding helicase (Kruisselbrink, Guryev et al. 2008). A similar observation was also made in a human cell line deficient of FANCJ, a homolog of DOG-1, with a large genomic deletion in the vicinity of G4 sequences (London, Barber et al. 2008).

Another study by Ribeyre and collaborators demonstrated the mutagenic potential of G4 structures by introducing a minisatellite CEB1 into the genome of the genetically tractable yeast *S. cerevisae*. The minisatellite CEB1 with 1.8 kb of tandem repeats, shown to form G4 *in vitro*, was markedly rearranged in PIF1-deficient cells, displaying large deletions and expansion of the repeat number, compared to the mutated non-G4 forming CEB1 control. Furthermore, the study proposed a mechanism leading to CEB1 instability, suggesting that the rearrangement occurs through repairing recombinogenic structures, possible G4 induced double-strand breaks in this case, via homology-dependent strand displacement and annealing (Ribeyre, Lopes et al. 2009). In support of this, the instability is shown to be associated with the formation of Rad51-Rad52-dependent recombination intermediates (Ribeyre, Lopes et al. 2009) (Lopes, Piazza et al. 2011).

1.4.1.7.7 Roles of G4s in maintaining telomeres

In most telomeric DNA, guanines and cytosines are distributed asymmetrically between the two strands: guanines are enriched on the strand with 5'->3' direction running from centromere to telomere. This strand is longer than its complementary strand, forming a G-rich single-stranded overhang at the very end of telomeres, which has been shown to form stable G4s *in vitro* (Bochman, Paeschke et al. 2012). It is thought that G-quadruplexes serve as protective structures to prevent the telomeric overhangs from being recognised by cellular surveillance mechanisms as broken DNA ends which could lead to checkpoint activation (Tarsounas and Tijsterman 2013). Supporting this idea, a recent study in yeast demonstrated that stabilising telomeric G-quadruplexes can suppress telomere capping defects in cells lacking CDC13, an essential protein for telomere capping (Smith, Chen et al. 2011).

Furthermore, many components of the shelterin complex, including RAP1, POT1 and TEBPβ have been shown to facilitate the assembly of G4s and bind to them (Giraldo and Rhodes 1994) (Paeschke, Juranek et al. 2008).

However, G4 formation during DNA replication represents a potential threat to telomere integrity. During this process, the telomeric G-rich overhang serves as the lagging strand, which is replicated discontinuously and hence transiently single-stranded, making opportunity for intramolecular G-quadruplexs to form *in vivo* (Bochman, Paeschke et al. 2012) (Schaffitzel, Berger et al. 2001). If not resolved by unwinding G4 helicases, the cells can suffer from severe telomere dysfunction and instability (Gomez, Wenner et al. 2006) (Vannier, Pavicic-Kaltenbrunner et al. 2012).

1.4.2 R-loops

1.4.2.1 Transcriptional-mediated formation of R-loops

RNA:DNA hybrids are common in replication and transcription processes. During the replication of the lagging strand, 11 nucleotide RNA primers generated by DNA primase complimentarily hybridise with the DNA template strand (Aguilera and Garcia-Muse 2012). RNA:DNA hybrids also occur between the nascent RNA transcript and DNA template before exiting RNA polII inside the transcription bubble, and these are 8 nucleotide in length (Westover, Bushnell et al. 2004). R-loops are formed and removed all the time but their accumulation is uncommon. An R-loop is a three-stranded structure consisting of a DNA:RNA hybrid and a displaced single-stranded DNA, identical to the RNA strand. So far most R-loops described *in vivo* are products of transcription (Aguilera and Garcia-Muse 2012). They are also key

intermediates in a prokaryotic immune system called CRISPR (Cluster Regularly Interspersed Short Palindromic Repeats). In this system, R-loops are established by base pairing of crRNA, a subcomponent of the protein- ribonucleic acid complex component called Cascade, to target DNA sequence (see review by (Ivancic-Bace, Howard et al. 2012). An R-loop can be longer than 1kb (formed in Immunoglobulin Class Switch Region) (Yu, Chedin et al. 2003).

Two mechanisms have been proposed for the formation of R-loops. In the first mechanism, R-loops occur as an extension of the usual RNA:DNA hybrid formed in the transcription bubble as RNA polII elongates. However, this model is strongly inconsistent with crystallographic studies showing that RNA transcript and DNA template exit through two different channels. The more accepted model suggests that R-loops arise when the nascent RNA transcript "threads back" as soon as it exits from RNA polII and hybridises to the template strand before the DNA duplex re-anneals (Westover, Bushnell et al. 2004).

Although mostly described in RNA polII transcription, R-loops have also been found over ribosomal DNA during transcription by RNA polI (El Hage, French et al. 2010). A recent study provides evidence for the presence of R-loops in RNA polIII transcribed tRNA genes in *S. cerevisiae* (Chan, Aristizabal et al. 2014).

R-loop formation has been extensively described as an *in cis* event (i.e co-transcriptionally and by the invasion of a nascent RNA transcript into the DNA duplex); however recently *trans*-induced R-loops have been discovered as a consequence of Rad51p and Rad52p-dependent strand exchange in homologous

recombination. Using a YAC containing a homologous sequence with a GALpr region, Wabba *et al*. show that activation of transcription of the endogenous GALpr induces R-loop formation on the YAC and YAC instability, and that the level of YAC R-loop is greatly reduced in the absence of Rad51p and Rad52p (Wahba, Gore et al. 2013).

1.4.2.2 Properties and features of R-loops

RNA:DNA hybrids are more thermodynamically stable than double-stranded DNA (dsDNA) (Roberts and Crothers 1992). One explanation for R-loop stability is that RNA:DNA hybrid adopt a conformation that is intermediate between molecules consisting of all DNA or all RNA (Shaw and Arya 2008). Another possibility is that the displaced DNA strand can adopt G-quadruplexes which stabilise the R-loop (Duquette, Handa et al. 2004).

Many studies have shown that R-loops are preferentially formed when the non-template DNA strand is G-rich (Yu, Chedin et al. 2003, Li and Manley 2005, Ginno, Lott et al. 2012, Powell, Coulson et al. 2013). Moreover, work by the Lieber's laboratory shows that the formation of R-loops also depends on other factors such as G:C content, DNA supercoiling and DNA nicks. According to their studies, the presence of guanine clusters on the non-template strand is extremely important in initiating R-loops formation (R-loop Initiation Zone-RIZ), but once initiated R-loop elongation does not rely on G clusters and is primarily determined by the density of guanine (R-loop Elongation Zone-REZ) (Roy and Lieber 2009). Furthermore, negative supercoiling facilitates R-loop formation by favouring breathing of two DNA

strands and reducing the G dependency. In addition, DNA nicks in the non-template strand downstream of the promoter reduce the chance of reannealing with the template strand, promoting the formation of the DNA:RNA hybrid (Roy, Zhang et al. 2010).

1.4.2.3 Regulation of R-loops

Cells have developed a number of pathways to remove R-loops. The most characterised one is RNase H enzymes, which digest the RNA strand of the hybrid in a sequence-independent manner. There are two classes of RNase H present in the most organisms, type 1 (monomeric protein complex) and type 2 (heteromeric complex) (see review by (Cerritelli and Crouch 2009)). The other R-loop removal pathways involves helicases that resolve R-loops, such as RecG and RecA helicases in bacteria (Hong, Cadwell et al. 1995), Pif1 DNA helicase in yeast (Zhou, Zhang et al. 2014) (Boule and Zakian 2007), Sen1/senataxin (Mischo, Gomez-Gonzalez et al. 2011, Skourti-Stathaki, Proudfoot et al. 2011) and DHX9 RNA helicase in human (Chakraborty and Grosse 2011).

In addition, there are also mechanisms to prevent R-loop formation. Genome-wide studies reveal that disturbing RNA biogenesis and metabolism leads to an increase in R-loop level (Huertas and Aguilera 2003) (Li and Manley 2005) (Wahba, Amon et al. 2011, Stirling, Chan et al. 2012). In addition, topoisomerase I relaxes negative supercoiling that promotes opening of the DNA duplex, promoting DNA:RNA hybrid formation (Drolet, Phoenix et al. 1995) (El Hage, French et al. 2010) (Tuduri, Crabbe et al. 2009). Another R-loop regulation mechanism involves suppression of proteins

that promotes DNA:RNA hybrids such as AtNDX and Rad51p (Sun, Csorba et al. 2013) (Wahba, Gore et al. 2013).

1.4.2.4 Detection of R-loops

One challenge in R-loop studies is the ability to detect R-loops *in vivo*. Capturing R-loops in the genome can be achieved by immunoprecipitation using the monoclonal antibody S9.6 which specifically recognises DNA:RNA hybrids (Boguslawski, Smith et al. 1986) (El Hage, French et al. 2010) (Mischo, Gomez-Gonzalez et al. 2011) (Ginno, Lott et al. 2012), or by inactivated RNase H (Ginno, Lott et al. 2012). Imaging R-loops using electron microscopy (EM) allows direct visualisation of R-loops (Duquette, Handa et al. 2004). Immunofluorescence with the antibody S9.6 is also used to detect locations of DNA:RNA hybrids (Powell, Coulson et al. 2013). Finally, R-loops can be indirectly detected by non-denaturing treatment of sodium bisulfite. This chemical converts cytosines on the single-stranded DNA of an R-loop to uracils. Sequencing of the treated DNA allows estimation of the length of the R-loop (Gomez-Gonzalez and Aguilera 2007).

1.4.2.5 Biological functions of R-loops

1.4.2.5.1 Roles of R-loops in DNA replication and recombination

R-loops have been described as an alternative DNA replication mechanism in bacteriophage T4 and ColE1-type plasmid in *E. coli* (Itoh and Tomizawa 1980) (Kreuzer and Brister 2010). Unlike the initiation of DNA replication in the

chromosome which relies on oriC sequence, ColE1 plasmid initiation requires RNA polymerase (RNAP) and is dependent on a RNAP-driven sequence, called RNAII, which forms a stable DNA:RNA hybrid with the template strand. DNA replication is started when this RNAII is cleaved by RNase H1 to create a group of 3' ends which is extended by DNA polymerase (Itoh and Tomizawa 1980). A similar phenomenon has also been found in eukaryotic mitochrondrial DNA, which has a similar structure as bacterial plasmids (Xu and Clayton 1996).

In addition, R-loop formation has been shown as an essential event in the mechanism and regulation of immunoglobulin class switch recombination (CSR). CSR occurs at the switch region (S) which is a several kb region of 25-80bp repeats and is G-rich on the non-template strand (Gritzmacher 1989) (Dunnick, Hertz et al. 1993). RNA:DNA hybrids have been shown to form during transcription of S regions *in vitro* and in B cells (Reaban and Griffin 1990) (Yu, Chedin et al. 2003). The DNA single strand of the hybrids is targeted by B cell-specific cytidine deaminase AID, which leads to double-strand breaks responsible for CSR (Chaudhuri and Alt 2004). Studies have shown that the formation of R-loops at the S regions only occurs in the physiological orientation with G-richness being on the non-template strand (Reaban and Griffin 1990) (Daniels and Lieber 1995). Inverting this region, so that the non-template strand is C-rich, or reducing the G content results in a dramatic reduction of CSR (Chaudhuri and Alt 2004).

1.4.2.5.2 R-loops and DNA methylation

R-loops have been found to be involved in DNA methylation in the Chedin laboratory. They first showed that the majority of strong CpG island promoters are characterised by positive GC-skew downstream of transcription, with the non-template strand displaying an excess of G over C residues, a condition that favours R-loop formation. Using non-denaturing bisulfite treatment, promoter CpG islands of human imprinted gene SNRPN, non-imprinted gene APOE and murine locus Airn are shown to adopt R-loop structures. Finally, to investigate the association of R-loops and the DNA methylation status of these loci, the CpG islands were cloned into an episomal system in both orientations (non-template strand is G-rich or C-rich) and then used as substrates for intracellularly expressed DNMT3B and/or DNMT3L. The results show that only in the orientation permissive for R-loop formation, are the CpG islands protected from DNA methylation by the enzymes, suggesting that R-loops prevent DNA methylation. However, the mechanism of this phenomenon is yet to be uncovered (Ginno, Lott et al. 2012).

1.4.2.5.3 Roles of R-loops in transcription

Regulation of gene expression usually involves elements at the 5' and 3' end of the gene. Using genome-wide approaches, Ginno *et al*. demonstrate that R-loops can form at CpG island promoters with a strong positive GC skew, which associates with unmethylated status (see above) and H3K4me3 enrichment (Ginno, Lott et al. 2012) (Ginno, Lim et al. 2013). This links R-loop formation with gene activation or maintenance of gene activation.

Several groups have documented the prevalence of G-rich sequences at the 3' end of genes using genome-wide analysis and bioinformatic approaches. Salisbury *et al*. noted that G-rich sequences are a relatively common feature immediately downstream of poly(A) signals in mammalian cells (Salisbury, Hutchison et al. 2006). Overrepresentation of G4 motifs was also observed immediately after the 3' ends of genes by Huppert *et al*., especially in those cases where the second gene is in close proximity, posing a possibility that G4 might be involved in transcription termination (Huppert, Bugaut et al. 2008). Moreover, analyses by Ginno *et al*. suggest that a group of 3' end with positive GC skew is enriched with R-loops (Ginno, Lim et al. 2013). Whether these regions that form R-loops also adopt G4s is yet to be investigated.

The speculation that R-loops might play a role in transcription termination has been investigated by the Proudfoot laboratory, which has led to a discovery of a new function of R-loops. It is known that the transcription termination process depends on both a functional poly(A) signal and downstream terminator sequences (Connelly and Manley 1988) (Whitelaw and Proudfoot 1986, West, Proudfoot et al. 2008). There are two classes of terminator sequences: CoTC RNA sequences and transcription pause sites. Skourti-Stathaki *et al*. showcase R-loop formation over the G-rich transcription pause site of β-actin and the MAZ4 pause sequence cloned downstream of a β-globin poly(A) signal in HeLa cells. Depletion of the RNA/DNA helicase senataxin leads to aberrant transcriptional termination, represented by an increase in readthrough transcripts *in vitro* and *in vivo*. A similar effect is observed when RNase H is overexpressed in the cells, suggesting both formation and resolution of R-loops are important for efficient termination. Furthermore, senataxin is required for the

recruitment of 5'-3' exonuclease Xrn2 that is necessary for termination. Based on this data, they proposed a model in which R-loop formation over a transcription termination site is required to pause PolII downstream of the poly(A) signals, but then the loops have to be resolved by senataxin to allow Xrn2-mediated degradation of the nascent transcript from the site of poly(A) cleavage, resulting in termination (Skourti-Stathaki, Proudfoot et al. 2011). However, it remains unclear how the cells achieve such a fine balance between the formation and resolution of R-loops in this case.

1.4.2.5.4 Threats to genome stability

The single DNA strand in R-loops could be very problematic and potentially poses a great danger to genome integrity. A DNA duplex is protected from DNA damage agents that are present in the cells because the two strands base pair with each other and are wrapped around proteins (histones). In R-loops, because the displaced ssDNA has lost both of these shields, it is susceptible to lesions, mutagenesis (Transcription-associated mutagenesis –TAM) or recombination (Transcription-associated recombination-TAR). One possible mechanism leading from ssDNA to DNA damage and mutagenesis is deamination. This process converts cytosine to uracil and is catalysed by activation-induced cytidine deaminase (AID) enzyme (Muramatsu, Kinoshita et al. 2000) (Revy, Muto et al. 2000). The resulting mismatch U:G is then replicated and gives rise to a mutation. Alternatively, the uracil could be processed by base excision repair (BER) components such as uracil–DNA glycosylase and abasic endonuclease (APE), both of which could result in DNA nicks or mutation. DNA nicks could further lead to double strand breaks (Conticello 2008) (Hamperl and Cimprich 2014) (Skourti-Stathaki and Proudfoot 2014).

Furthermore, there are other features of R-loops that are targeted by enzymes that cleave the structures, resulting in single strand breaks. For example, if G-quadruplexes are formed on the transiently displaced ssDNA, they can attract GQN1 nuclease that cuts within the single-stranded region 5' of the G quartet stack (Sun, Yabuki et al. 2001). In addition, each R-loop has two duplex-single strand junctions, which are recognised and cleaved by several structure-specific endonucleases such as XPF-ERCC1 and XPG (Tian and Alt 2000).

Another mechanism that could explain the catastrophic effects of R-loops on genome stability is the collision with on-going replication forks. This was first predicted by the observation that R-loops are able to impair transcription elongation, which in turn blocks the progression of DNA replication. R-loops adopted during transcription have been shown to impede replication forks in a bacterial plasmid carrying a Ig-class switch region by 2D gel electrophoresis of replication intermediates and in mammalian cells by fiber analysis of on-going replication forks (Gan, Guan et al. 2011). One example of such a block is observed in chicken DT40 and HeLa cells deficient for the splicing factor SRSF1, causing DSBs (Li and Manley 2005) (Gan, Guan et al. 2011). Transcription machinery and replication forks can collide when DNA and RNA polymerases move in the same direction (co-directional collision) but with different velocities or when they move towards each other (head-on collision). Such a replication impairment could generate DNA lesions and double-strand breaks, which induce recombination-mediated repair that in turn leads to rearrangement and genome instability (Aguilera and Garcia-Muse 2012) (Hamperl and Cimprich 2014).

Deleterious effects of R-loops on genome instability have recently been linked to an accumulation of the chromatin condensation mark, histone H3S10 phosphorylation (H3S10P) (Castellano-Pozo, Santos-Pereira et al. 2013). It has been found that R-loop accumulation in *C.elegans*, yeast and human cells depleted of THO/TREX or SETX coincides with an overabundance of H3S10P at pericentric heterochromatin, centromeres and other active regions of the genome (e.g open reading frames), and this is suppressed by RNase H overexpression (Dominguez-Sanchez, Barroso et al. 2011) (Skourti-Stathaki, Proudfoot et al. 2011) (Castellano-Pozo, Santos-Pereira et al. 2013). Such compaction, as a consequence of an increase in the H3S10P mark, would contribute to replication fork stalling, transcription impairment and silencing (Castellano-Pozo, Santos-Pereira et al. 2013) (Skourti-Stathaki and Proudfoot 2013).

1.4.2.5.5 R-loop regulation of non-coding RNA

A growing body of evidence suggests the roles of R-loops in gene regulation through interfering with the expression of non-coding RNA. The first example can be taken from the study by Sun *et al*. on the regulation mechanism of the Arabidopsis flowering locus (FLC). The expression of this locus is modulated by the antisense generated from another locus called COOLAIR. Sun and colleagues found that R-loops can be formed over the G-rich region of the COOLAIR promoter and if they are stabilised by a protein called AtNDX, which binds single-stranded DNA, COOLAIR antisense expression will be inhibited, which in turn modifies FLC expression (Sun, Csorba et al. 2013).

The second example is the regulation of the *Ube3a* imprinted gene, mutations of which cause Angelman syndrome (AS). The expression of this gene is controlled by *Ube3a* antisense (*Ube3a-ATS*) located immediately downstream of the *Snord116* gene. Normally, Ube3a antisense transcripts silence the Ube3a gene in *cis*. Powell and colleagues demonstrated that R-loop formation occurs over a repetitive intronic region of high GC skew at *Snord116*. Furthermore, they showed that Topotecan, a topoisomerase I inhibitor, treatment increases R-loop formation that leads to chromatin decondensation and stalled transcriptional progression through *Ube3a-ATS*, which results in the loss of the antisense transcripts and hence derepression of the *Ube3a* gene (Powell, Coulson et al. 2013).

1.4.2.5.6 R-loops in diseases

The mechanistic link between R-loop formation and neurodegenerative diseases has recently been documented. In addition to the example above about Angelman syndrome, a disorder associated with epilepsy and a severe learning deficit, R-loops have been demonstrated to trigger gene silencing in Friedreich Ataxia, Fragile X syndrome, and Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). These conditions all have a common feature in that they are caused by an expansion of repeated DNA sequences: GAA in the first intron of FXN gene in Friedreich Ataxia, CGG in FMR1 gene in Fragile X syndrome and GGGGCC in C9orf72 gene in ALS/FTD (Loomis, Sanz et al. 2014) (Colak, Zaninovic et al. 2014) (Groh, Lufino et al. 2014) (Haeusler, Donnelly et al. 2014). R-loop formation over these expanded repeats mediate gene silencing by promoting the deposition of H3K9me3 repressive mark in Friedreich Ataxia and Fragile X syndrome (Groh,

Lufino et al. 2014) or by impeding transcriptional progression resulting in an accumulation of abortive transcripts in ALS/FTD (Haeusler, Donnelly et al. 2014).

1.5 THESIS AIMS

As discussed in part 1.3.2.4, it has been demonstrated that the chromatin remodeler ATRX localises at various locations in the genome. While it is clear that H3K9me3 is the targeting signal for ATRX localisation at the A/T rich pericentric heterochromatin, the targeting mechanisms for localising ATRX to the vast majority GC-rich sites, including telomeres, rDNA repeats, GC-rich tandem repeats and CpG islands, are unknown. The aim of this study is to determine the targeting signal for ATRX localisation at the GC-rich tandem repeats.

There are two observations made in the previous study that are crucial for the rationale of the experimental design for this study. The first observation is the correlation between the size of the ψζ G-rich tandem repeat in the alpha-globin cluster and the reduction of alpha-globin gene expression in the absence of ATRX (Law, Lower et al. 2010). Given the fact that ATRX directly binds to this repeat as determined by ChIP, how might the repeat size influence the recruitment of ATRX? Answering this question will help us understand the enigma of the alpha-thalassemia variations in ATRX syndrome patients. The second observation is that some ATRX targets are tissue-specific (Law, Lower et al. 2010), suggesting that ATRX targeting might be linked to transcription activity. This prompts a question, how does transcription actually influence the ATRX recruitment? Finally, GC-rich repeats are prone to form non-canonical DNA structures, two of which, the G-quadruplexes and

R-loops, have been studied the most intensively. How are these factors, the G-rich repeat size and transcriptional activity related to the propensity for DNA secondary structure formation and do these contribute to the recruitment of ATRX to its targets?

Chapter 3 of this thesis will describe the establishment of an experimental model where all the factors discussed above are integrated to investigate ATRX binding at GC-rich tandem repeats. Next, in chapter 4, chromatin immunoprecipitation is used to examine ATRX recruitment to the repeats with variations in the sequences, size, transcription and the orientation of transcription. Lastly, in chapter 5 I will explore possible features of the repeats that can serve as ATRX targeting signal at the repeats with the focus on G-quadruplexes and R-loops.

2 Materials and Methods

2.1 MATERIALS

Table 2-1 Primer oligos. All primers oligos are synthesized by Thermo Scientific.

PRIMER NAME	SEQUENCE (5'-3')	PURPOSE
ψζVNTR F	CAGTCACCTGAGGCAAGACCTACTTC CCGCACTTCGAC	PCR
ψζVNTR R	CAGTCAACGCGTTCCTTCCCTCCAAT ACTGAGAACAC	PCR
Upstream_Int_F	AATAGCCTCCTGACCACAGATCCTT	Sequencing
Downstream_Int_R	CCCCATGAGAACCCACAGTGTT	Sequencing
DownstreamVNTR1 F	GCGATCTGGGCTCTGTGTTC	ChIP Q-PCR
DownstreamVNTR1 R	TCTTTGCTGTGGAGTGGTATTCC	ChIP Q-PCR
DwnstmVNTR3 QPCR F	AGGGAAGGAACGCGCAAT	ChIP Q-PCR
DwnstrmVNTR3 QPCR R	TTTGCTGTGGAGTGGTATTCCA	ChIP Q-PCR
DIST F	GAGATGCTGGAGTCAGGACCAT	ChIP Q-PCR
DIST R	AGGAGTCAGGAGCAGCAGTCA	ChIP Q-PCR
GAPDH F	ACCTGTGCTCCCACTCCTGATTTC	ChIP Q-PCR
GAPDH R	TGCCAAGTTGCCTGTCCTTCC	ChIP Q-PCR
16p Telomere F	CCTCGCCTTGCCTTGGGAG	ChIP Q-PCR
16p Telomere R	CGGTTCAGTGTGGAAAATGGGAAAC	ChIP Q-PCR
Endogenous ψζRepeat F	CCACAGTTCCTGCCCTGACTCC	ChIP Q-PCR
Endogenous ψζRepeat R	TCTCCCCTCCTTCCCTCCAATAC	ChIP Q-PCR
NME4 Repeat F	TCCCTCCATCGGCTCTGAAAAG	ChIP Q-PCR
NME4 Repeat R	GCTCTGCTGCGGCTGGTGC	ChIP Q-PCR
rDNA p4 F	CCGTCCGTCCTTCCGTTCG	ChIP Q-PCR
rDNA p4 R	CCACCCCACCACCACGCC	ChIP Q-PCR
D1 F	AGACTGGAGCAACCCATTATCTAGA	ChIP Q-PCR
D1 R	AGTAGACTCGGGCTATTCAGTT	ChIP Q-PCR
U1 F	TGTTTTGTGGGATTGGTTGATAGT	ChIP Q-PCR
U1 R	TGAGCAAATGGGAAAAGTATTTCA	ChIP Q-PCR
U2 F	GGAATCTGTAGCCCACTGGATT	ChIP Q-PCR
U2 R	CACAGCTGTTCCCGACACAT	ChIP Q-PCR
rU2 F (for the reverse telomeric repeat)	CACAGCTGTTCCCGACACAT	ChIP Q-PCR
rU2 R (for the reverse telomeric repeat)	CGAAGCTAGCAATTGCCTAGGA	ChIP Q-PCR

D2 F	GTCCCCTAGTCCCTCACTGAA	ChIP Q-PCR
D2 R	TGCTCAGTCCAATCAGTGAGA	ChIP Q-PCR
D3 F	AATGCAGGTGAAGTTCGAGGG	ChIP Q-PCR
D3 R	CGTCCTCCTTGAAGTCGATGC	ChIP Q-PCR
D4 F	CTAGAGGGCCCGTTTAAACCC	ChIP Q-PCR
D4 R	CAAACAACAGATGGCTGGCAA	ChIP Q-PCR
D5 F	CGCTCCTTTCGCTTTCTTCC	ChIP Q-PCR
D5 R	CGATTTAGAGCTTGACGGGGA	ChIP Q-PCR
pSV40 F	CCCCATGGCTGACTAATTTT	to check the integration of the cassette into genome
5' Hygromycin R	CCTCCTACATCGAAGCTGAA	
GAPDH promoter F1	TATAAATTGAGCCCGCAGCCT	ChIP Q-PCR
GAPDH promoter R1	AAAGAAGATGCGGCTGACTGT	ChIP Q-PCR
GFP promoter F	TGACTCACGGGGATTTCCAAG	ChIP Q-PCR
GFP promoter R	CCCGTTGATTTTGGTGCCAAA	ChIP Q-PCR
Human HBB Promoter F	AACTGTGTTCACTAGCAACCTCAAA	ChIP Q-PCR
Human HBB Promoter R	ACAGGGCAGTAACGGCAGACT	ChIP Q-PCR
ZNF180 F	TGATGCACAATAAGTCGAGCA	ChIP Q-PCR
ZNF180 R	TGCAGTCAATGTGGGAAGTC	ChIP Q-PCR
TEL-REPT ORI1 F	TAGGAAGGTGGATCACCTGAGGTGC ACGCCCAGATCACGGATGTGTC	to amplify telomeric repeat of the original orientation from P. Thornton patient
TEL-REPT ORI1 R	TCGAAGCTAGCAATTGCGCGCCCTAA CCCTAACCCTAACCCTAACCCTA	
GFP Exon2 F	GTGAACCGCATCGAGCTGAAG	Southern Blot probe
GFP Exon2 R	CTTGTACAGCTCGTCCATGCC	
GFP QPCR Fwd	ACGACGGCAACTACAAGACC	QPCR mature GFP mRNA
GFP QPCR Rev	GTCCTCCTTGAAGTCGATGC	

Plasmid vector

pcDNA5/FRT/GFP GAA plasmid is a gift from Dr. Marek Napierala (Texas, USA).

pOG44 plasmid are purchased from Invitrogen.

RNase H1-GFP vector plasmid is kindly provided by the Proudfoot laboratory (Sir William Dunn School of Pathology, Oxford).

Genomic DNA

Human genomic DNA of healthy individuals is obtained from anonymous donors. The DNA was extracted by Helena Ayyub.

Cell line

Flp-In™ 293 T-Rex cell line (R780-07) is purchased from Invitrogen.

E. coli strain

Stbl2 strain for cloning unstable and repetitive DNA, and DH5α for propagating other types of DNA are obtained from Invitrogen.

Antibodies

Table 2-2 Antibodies used in the study.

ATRX	Rabbit polyclonal H300 (sc-15408) 200μg/mL, Santa Cruz. Used 1: 100 dilution for Western Blot, 10μg per ChIP reaction.
RNA polII	Rabbit polyclonal N20 (sc-899), Santa Cruz. Used at 10μg per ChIP reaction.
H3K9me3	Rabbit polyclonal antibody (Ab8898), Abcam. Used at 5μg per ChIP reaction.
Rabbit IgG	Negative control rabbit IgG (X0903), Dako. Used at 10μg per ChIP reaction.
Mouse IgG	Normal mouse IgG (sc-2025), Santa Cruz. Used at 10μg per ChIP reaction.
G4	G4 antibody (1H6). Kindly provided by the Lansdorp laboratory. Used at 10μg per ChIP reaction.
R-loops	Mouse monoclonal antibody S9.6 (HB-8730), ATCC. The hybridoma is cultured in the lab and the cultured medium is obtained and used at 10μL per ChIP.

2.2 METHODS

2.2.1 PCR amplification and sequencing G-rich tandem repeats

The ψζVNTR from human DNA was amplified using primers ψζVNTR F and R with sequences in Table 2-1. The reaction was carried out with addition of DMSO buffer (16.6mM (NH4)2SO4, 67mM Tris-HCl, 10%DMSO and 10mM Dimecarptoethnol) and 5M Betaine to overcome the high GC content and highly repetitiveness of the template. Thermal cycles were as follows: 94^{o}C for 2 minutes, 35 cycles of 94^{o}C for 1 minute, 55^{o}C for 30 seconds, 72^{o}C for 1 minute 30 seconds, and 72^{o}C for 10 minutes.

The human telomere repeats were amplified from the genomic DNA of patients that have a truncated chromosome fused with telomere sequences the primer pair TEL-REPT ORI1 F/R (Table 2-1). The PCR reaction was performed with Faststart Taq polymerase and the thermal cycles as follows: 95^{o}C for 5 minutes, 35 cycles of 95^{o}C for 30 seconds, 57.7^{o}C for 30 seconds, 72^{o}C for 1 minute 30 seconds, and 72^{o}C for 10 minutes.

To sequence the G-rich tandem repeats, a sequencing PCR reaction was carried out using 200-300ng of template DNA, 4µL of Big Dye (Applied Biosystem), 4µL of Big Dye Terminator buffer (Applied Biosystem), 4µL of 5M Betaine and 1µL of 3.2µM primer. The thermal cycles were as follows: 96^{o}C for 15 seconds, 25 cycles of 96^{o}C for 45 seconds, 50^{o}C for 45 seconds and 60^{o}C for 3 minutes 30 seconds. Following the reaction, the sequencing PCR products were precipitated in 95% Ethanol for 10 minutes and then cleaned up using 70% Ethanol. The purified DNA was then resuspended in 16µL Hi-Di Formamide (Life Technologies) and sent to the

sequencing facilities (Weatherall Institute of Molecular Medicine) for sequence reading.

2.2.2 Plasmid Construction

The pcDNA5/FRT/GFP GAA plasmid has the backbone of the pcDNA5/FRT/TO (Invitrogen) and GFP gene in the multi-cloning site. This GFP gene has two exons and exogenous intron (intron of Rat Pem gene) with a stretch of GAA repeats flanked by Bsu36I and BssHII recognition sequences. Its expression is driven by a CMV promoter, which is under regulation of two consecutive tetracycline operators. The vector also has a homologous recombination site, FRT, which can be used to integrate the vector into the genome. The PCR products of ψζVNTR are double-digested with Bsu36I and BssHII, then replaced GAA repeat stretch on pcDNA5/FRT/GFP GAA vector to make pcDNA/FRT/GFP ψζVNTR vector.

To construct the No VNTR control vector, the (GCGGG) repeat stretch of ψζVNTR is removed from the pcDNA/FRT/GFP ψζVNTR using BssHII restriction enzyme whose recognition site flanks the entire stretch.

2.2.3 Colony lift assay

This technique is used for high through-put screening for bacterial colonies that have harbored a recombinant DNA fragment of desire.

Preparing the probe. 50ng of probe DNA was placed into a microcentrifuge tube with 5µL of primers and 30µL of water to an approximate total volume of 35 µL. The probe and primers were denatured by heating to 95^{o}C for 5 minutes. 10µL of Labelling buffer, 2µL of alpha-dCTP and 2µL of Klenow fragment were added and incubated at 37^{o}C for 10 minutes. Un-incorporated alpha-dCTP was removed by filtering the reaction through a Sephadex G50 column and centrifuging at 735G for 1 minute. The labeled probes were denatured at 95^{o}C for 5 minutes before adding to PerfectHyb Hybridization buffer (Sigma Aldrich, H7033).

Preparing the plates and filters. Hybond filters (GE Health Care) were labeled and then dropped and patted on screening bacterial plates. The orientation of the plates and the filters was denoted by piercing an inked needle through the filters. The filters were then peeled off and placed with the colonies side up onto a piece of Whatman paper soaked in 2xSSC/5% SDS in for 2 minutes. The filters were heated in the microwave on full power for 5 minutes to denature the DNA, and then were put into a hybridization chamber with PerfectHyb Hybridization Buffer (Sigma Aldrich, H7033) containing the probe. The hybridization was done at 65^{o}C for 2 hours.

Washes and Autoradiography. After 2 hours, the filters were washed once with cold wash buffer (2X SSC) for 10 minutes and once with hot wash buffer (0.1X SSC, 0.1% SDS), which was pre-heated to 65^{o}C, for 5 minutes. Autoradiography films were applied onto the filters and then exposed after 1 hour. The exposed films were then aligned with the ink marks on the filters.

2.2.4 Cell culture

Flp-In T-REx-293 (293TRex) cell line contains two stable, independently integrated plasmids which exhibit the following features. The pFRT/lacZeo plasmid introduces a single FRT site into the genome and stably expresses the LacZ-Zeocin fusion gene. The location of the FRT site has not been mapped but presumably in a transcriptionally permissive genome locus, as determined by generation of a Flp-In™ expression cell line containing the pcDNA5/FRT/CAT control plasmid. The pcDNA6/TR plasmid stably expresses the tetracycline repressor gene and blascitidin resistance gene. 293TREx cells were cultured in DMEM (Invitrogen) supplemented with 10% fetal calf serum, 1% L-Glutamine and 1% Peni-Streptomycin. Blasticidin and zeocin were added in the medium with the concentrations of 15ug/mL and 100ug/mL, respectively, in order to maintain the FRT site in the genome and the expression of tetracycline repressor gene.

2.2.5 Introduction of ectopic ψξVNTR into 293TRex cells via Flp Recombinase-Mediated Integration

293TRex cells were cultured in 6 well plates until a confluence of more than 90% was achieved. The cells were then transfected with plasmid constructs to generate stable expression cell lines. Two mixtures were prepared for the transfection. Mixture A contained 20uL of Lipofectamin A (Invitrogen) and 230uL OptiMem I (GiBCO). Mixture B contained 1ug of plasmid pcDNA/FRT/GFP ψζVNTR, 9ug of plasmid pOG44, which produces Flp recombinase enzyme, and 250 uL of OptiMem I. Mixture A was incubated at room temperature for 10 minutes, then was combined with

mixture B. This combined solution was incubated at room temperature for 20 minutes. Prior to the transfection, the old medium in the plate was aspirated and 1.5mL medium (DMEM 20% FCS 1% L-Glutamine without Pen-Strep) was added to every well. Then the combined solution was overlaid onto the cells. 24 hours after transfection, the cells were trypsinised and plated into a fresh medium (without antibiotics) at a confluency of about 25%. The next day, a selective medium (DMEM 10% FCS 1% L-Glutamine 1% Pen-Strep) containing 200ug/mL hygromcyin B and 15ug/mL blasticidin was added to the plate. The cells were fed with the selective medium every 3-4 days until foci were identified. The hygromycin-resistant foci were picked and expanded. To verify that the pcDNA/FRT/GFP ψζVNTR constructs have integrated into FRT site, the cell clones were further tested for zeocin sensitivity and lack of β-galactosidase activity.

2.2.6 Zeocin sensitivity assay

The cells were plated in a medium containing hygromycin and blasticidin so they were at 25% confluent. The next day, the medium was replaced with a fresh medium containing 100ug/mL zeocin. The selective medium was replenished every 3-4 days for 2 weeks. Successfully integrated cells should be sensitive to zeocin. The zeocin sensitive cells first experience morphological changes such as a vast increase in size, a presence of large empty vesicles in the cytoplasm and an abnormal cell shape. These cells will break down with only strings of protein remaining.

2.2.7 Beta-gal staining

The cells were cultured so that they were 60-70% confluent on a 60mm plate. The medium was removed from the plate and the cells were rinsed with 1X PBS. After that, they were fixed with 3mL 1X Fixative Solution (Invitrogen) and then rinsed twice in 2.5mL 1X PBS. The cells were stained with Staining solution (20mg/mL X-gal, 400mM potassium ferricyanide, 400mM potassium ferrocynide and 200mM magnesium chloride) at 37oC for 2 hours. The successfully integrated cells will not turn blue because they lack β-galactosidase.

2.2.8 Induction of transcription

Selected clones were cultured in medium containing tetracycline free FBS (Takara Clontech), hygromycin (Invitrogen) and blasticidin (Invitrogen) until they reached a confluency of 45-50%. The medium was then removed and replaced with a fresh medium containing 1ug/mL doxycycline (which has the same mechanism of action as tetracycline) without the antibiotics. 24 or 48 hours after adding doxycycline, the cells were harvested for experiments.

2.2.9 Genomic DNA extraction

5-10 million cells were collected for DNA extraction. The cells were lysed by incubating with Lysis buffer (10 mM Tris pH8.0, 10 mM NaCl, 10mM EDTA and 0.5% SDS) at 37oC for at least 5 hours. DNA was precipitated by isopropanol and

washed in 70% ethanol. DNA pellets were collected by centrifuging at 13000rpm for 5 minutes, then air dried at room temperature and resuspended in water.

2.2.10 Southern Blot

15µg of genomic DNA of each clone was digested with KpnI or AflII restrictrion enzyme at 37oC for 5 hours. The digestion products were run on 1% agarose gel at 50V overnight. After that, the gel was soaked in 1M NaOH solution for 1 hour, then was neutralised for 2 hours in Neutralization buffer (1M Tris, 3M NaCl, 80mL/L concentrated HCl pH 7.4). DNA was blotted onto Zeta-probe nylon membrane (Bio-rad). The blot was then hybridized with GFP exon2 probe, which was labeled with ^{32}P at 42oC overnight. After probing, the membrane was washed with 2X SSC (0.03M Sodium citrate and 0.3M NaCl) and then with 0.1% SSC before developing onto a film.

2.2.11 Western Blot

$1x10^7$ cells of 293T-REx were lysed using high salt RIPA buffer (50mM Tris 7.4, 1M NaCl, 0.2% SDS, 1% Deoxycholic acid, 1% NP-40 and protease inhibitor cocktail). The lysate was incubated with 50µL of DNase I (Roche) for 20 minutes at room temperature. After that, cellular protein was collected by centrifugation at 13.000rpm for 10 min at 4oC. The concentration of whole cell protein was measured using a DC protein assay kit (Bio-rad). For western blot, 50µg of protein was used, as well as the same amount of positive control (+/+ATRX cells). The samples were denatured and then run on NuPAGE 4-12% Bis-Tris gel at 150V. Then, the protein on the gel was transferred onto a PVDF Immobilon-P membrane. The membrane was then incubated

with 39f monoclonal anti-ATRX antibody at 1:10 dilution, washed and developed onto a film.

2.2.12 Chromatin Immunoprecipitation (ChIP)

2.2.12.1 ATRX ChIP

ATRX chromatin immunoprecipitation is used to capture genome regions that are bound by ATRX, using an antibody against ATRX. The method is previously published (Law et al, 2010).

Cross-linking. Approximately $3x10^7$-$5x10^7$ cells were used for each ChIP. Cells were harvested by trypsinisation, then washed in PBS twice. The cells were then fixed with 2mM EGS at room temperature for 45 minutes in 20mL PBS. Formaldehyde was added to the fixing cell suspension with EGS to a final concentration of 1% for 20 minutes, then quenched with 0.125mM glycine.

Cell lysis. The cells were then lysed in a step-wise manner. First, cell membrane was lysed by Lysis Buffer 1 (100mM HEPES, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP40, and 0.25% Triton X-100). Cell nuclei were then washed by Lysis Buffer 2 (200mM NaCl, 1mM EDTA, 0.5mM EGTA and 10mM Tris pH8), then collected by centrifugation at 2500rpm for 2 minutes. And Lysis Buffer 3 (1mM EDTA, 0.5mM EGTA, 1mM Tris-HCl pH8, 100mM NaCl, 0.1% Sodium Deoxycholate and 0.5% N-lauroyl sarcosine) was used to lyse the nuclear membrane. All the lysis buffers were mixed with protease inhibitor cocktail (Roche complete Mini, EDTA-free).

Sonication. Chromatin was subjected to sonication in Lysis Buffer 3 for 30 minutes (30s ON, 30s OFF; High power, Diagenode Bioruptor) to yield DNA fragments of under 500bp. Cell debris was removed by centrifugation at 13000rpm at 40^{o}C, and the supernatant containing chromatin was obtained.

Preclearing and IP. The chromatin was pre-cleared with Protein A Dynabeads (Life Technologies, 10002D), which were pre-blocked with 0.5% BSA, for 1 hour. 50uL of the sonicated chromatin was taken as an input. The remaining chromatin was immunoprecipitated with Dynabeads pre-coated with 10µg of ATRX antibody H300 or with 10µg of rabbit IgG control over night at 4^{o}C.

Washes and Elution. The chromatin-antibody-bead complexes were washed five times with Wash Buffer (50mM HEPES, 1mM EDTA, 0.7% Sodium Deoxycholate, 500mM LiCl and protease inhibitor cocktail) and once with ice-cold PBS. All the washes were carried out for 5 minutes at 4^{o}C . The captured chromatin was eluted with 0.5% SDS and 0.1M Sodium Bicarbonate at room temperature.

Reverse cross-linking and purification. The crosslink was reversed with 0.2M Sodium Chloride at 65^{o}C for 4 hours. After that, protein K (Sigma Alrich) was added to break down the antibody-antigen bonds. DNA was purified and precipitated with ethanol, sodium acetate and carrier glycogen (Life Technologies).

2.2.12.2 Histone modification and RNA PolII ChIP

This technique was performed using reagents from ChIP Assay kit (Millipore, 17-295).

Cross-linking. Approximately $1x10^7$ cells were used for each ChIP. Cells were trypsinised and resuspended in PBS with cell density of $5x10^7$ cells/10mL. To cross-link the chromatin, Formaldehyde was added to the cell suspension at a final concentration of 1% and incubated for 10 minutes at room temperature. After that, 0.125mM glycine was added to quench the cross-linking. Cells were washed twice in cold PBS.

Cell lysis. The cells were lysed in SDS Lysis Buffer (mixed with protease inhibitor cocktail) for 10 minutes on ice.

Sonication. Chromatin was subjected to sonication for 12 minutes (30s ON, 30s OFF, High power, Diagenode Bioruptor). Cell debris was removed by centrifugation at 13000 rpm for 10 minutes at 4^{o}C. The sonicated chromatin was then diluted 1 in 10 in ChIP Dilution Buffer (mixed with protease inhibitor cocktail). 10uL of the diluted chromatin was taken as an input per antibody.

Pre-clearing and IP. 60uL of protein A/Salmon sperm DNA Agarose beads (from the kit) was added to pre-clear every 2mL of diluted chromatin for 1 hour at 4^{o}C. The diluted chromatin was then incubated with 10µg of H3K9me3 antibody, or of RNA polII antibody over night at 4^{o}C. An IP with normal rabbit IgG was used as a control. After that, 60uL of Agarose beads was added to the antibody-chromatin IP mix, and was incubated for 1 hour.

Washes and Elution. The antibody-chromatin-bead capturing complexes were washed once in Low Salt Buffer, High Salt Buffer, LiCl Buffer and twice in TE Buffer at 4°C. The captured chromatin was eluted in 1%SDS and 0.1M NaHCO3.

Reverse cross-linking and DNA purification. This step was performed as described above in ATRX ChIP.

2.2.13 Quantitative Polymerase Reaction Chain (Q-PCR)

Quantitative PCR was performed using SYBR green master mix (Applied biosystem 4309155). SYBR primers (Table 2-1) were designed manually or using MacVector software. All primers were tested by a five-point series dilution using genomic DNA. All Q-PCR reactions were performed in triplicates. ChIP or DIP enrichment was determined relative to input.

2.2.14 DNA Immunoprecipitation (DIP)

This technique is used to investigate R-loop enrichment, using a monoclonal antibody, S9.6, specific to RNA:DNA hybrids.

Cell lysis. Cells, cultured in a 10 cm plate to 70-80% confluency, untreated or treated with doxycycline as described above, were harvested by scraping in ice-cold PBS in a cold room (4°C). They were collected by centrifugation at 500g for 5 minutes at 4°C. The cells were then lysed by incubating with 600µL of Cell Lysis Buffer (0.5% NP-40, 85mM KCl and 5mM PIPES) on ice for 10 minutes. Nuclei were collected by

centrifugation at 2400rpm for 5 minutes at 4°C. Then 800µL of Nuclear Lysis Buffer (1%SDS, 25mM Tris-HCl pH8.0 and 5mM EDTA) was added and incubated on ice for another 10 minutes, followed by an incubation with 20 uL of Proteinase K (*Tritirachium album*, ≥800units/mL, Sigma Alrich) for 5 hours to overnight.

Genomic DNA extraction. Genomic DNA was isolated from the cell lysate by adding 400µL of gDNA isolation buffer (3M Potassium Acetate and 11.5% Glacial Acetic Acid). DNA in the supernatant after centrifugation at 13000rpm for 5 minutes at 4°C was collected, then precipitated with isopropanol and washed with 70% Ethanol. DNA pellets were air-dried and resuspended in water.

RNase H treatment. DNA of 50-100µg was divided into two halves. One was treated with 5µL recombinant RNase H (5U/µL, NEB) in a total volume of 100µL for 4 hours. The other half was used as a control. After the treatment, DNA was diluted with 300µL of IP Dilution Buffer (0.001% SDS, 1.1% Triton X-100, 0.12mM EDTA, 16mM Tris-HCl pH8.0, 160mM NaCl and protease inhibitor cocktail).

Sonication. DNA was sonicated for 15 minutes (30 sec ON, 30sec OFF) with High power setting by the Diagenode Bioruptor.

Pre-clearing and IP. Sonicated DNA was diluted in IP Dilution Buffer to a total volume of 3mL. 100µL of each sample was taken as an input. The remaining sonicated DNA was pre-cleared for 1 hour with Protein A Dynabeads, which were pre-blocked with 1% BSA in IP Dilution Buffer for 1 hour. The pre-cleared DNA was collected and incubated with either 10µg of S9.6 antibody or of normal mouse IgG

overnight at 4°C. The antibody-R-loops complexes were pulled down by incubating with 60µL Protein A Dynabeads pre-blocked with 1% BSA for 2 hours.

Wash and Elution. The beads were then collected and washed once with Buffer A (20mM Tris-HCl pH8.0, 2mM EDTA, 0.1% SDS, 1% Triton-X 100, 165mM NaCl and protease inhibitor cocktail), Buffer B (20mM Tris-HCl pH8.0, 2mM EDTA, 0.1% Triton-X 100, 0.55M NaCl and protease inhibitor cocktail) and Buffer C (100mM Tris-HCl pH8.0, 1mM EDTA, 1% NP-40, 0.01% Sodium Deoxycholate, 250mM LiCl and protease inhibitor cocktail). Each wash was done for 5 minutes at 4°C. After that, the beads were washed twice with Buffer D (100mM Tris-HCl pH8.0, 1mM EDTA and protease inhibitor cocktail) at 4°C and once at room temperature. Captured DNA was eluted twice in 250µL Elution Buffer (1% SDS and 0.1M NaHCO3).

DNA purification. Eluted DNA, together with the inputs, was purified using a QIAGEN PCR purification kit.

2.2.15 RNaseH-GFP transient transfection

One day before transfection, cells (of the stable inducible 293T-REx clones) were seeded in 6 well plates in the medium containing hygromycin and blasticidin so that they reached a confluency of 80% next day. Solution A was prepared with 15µL Lipofectamine 2000 and 150µL OptiMEM-I and incubated at room temperature for 5 minutes. Solution B was prepared by mixing 1µg DNA of RNase H1-GFP plasmid and 150µL OptiMEM-I. Then, solution A and B were mixed together and incubated at room temperature for 20 minutes. Prior to the transfection, the old medium in the

plate was aspirated and 1.5mL medium (DMEM 20% FCS 1% L-Glutamine without Pen-Strep) was added to the wells. Then the mixed solution was overlaid onto the cells. Following 4-5 hours of incubation at 37oC, the cells were transferred to a normal medium.

2.2.16 Fluorescence-Activated Cell Sorting (FACS)

This technique was used to check the expression of the ectopic GFP gene in the stable inducible expression 293T-REx clones and to check the expression of the RNase H1 transiently transfected into the clones.

Cells were harvested and washed once in PBS. They were then analysed on the CyAn Advanced Digital Processing (ADP) High-Performance Flow Cytometer.

3 Generation of stable expression mammalian cell line clones containing ectopic repetitive GC-rich tandem repeats

3.1 INTRODUCTION

As discussed in Chapter 1 Introduction, ATRX has been found to be highly enriched at GC-rich tandem repeats *in vivo*. However, it is unclear whether ATRX targeting is solely determined by the repeats themselves, by the chromatin environment, by other factors or by a combination of these. This question can be investigated by placing a repeat ectopically at a random location in the genome and ask if it recruits ATRX.

To this end an experimental model was used, in which a known G-rich tandem repeat target of ATRX is ectopically introduced into an ATRX wild-type cell line. Repeats were placed in an intron of a reporter gene (GFP) which is regulated by an inducible promoter and allows gene expression to be switched on and off (Figure 3.1).

The Flp-In T-REx system (Life Technologies) is ideal for the purpose of this study for several reasons. First of all, the composition of the insert sequence could be manipulated. Secondly, the incorporation of the ectopic construct was performed via Flp recombinase-mediated site-directed recombination at a single specific site (FRT) (Figure 3.2), hence controlling for the genomic context. Since all of the integrated clones are isogenic in terms of chromatin context around the ectopic repeat, comparison between clones carrying different features of the ectopic repeat is

appropriate. Any difference in ATRX binding between different clones will be attributed only to the differences at the repeat.

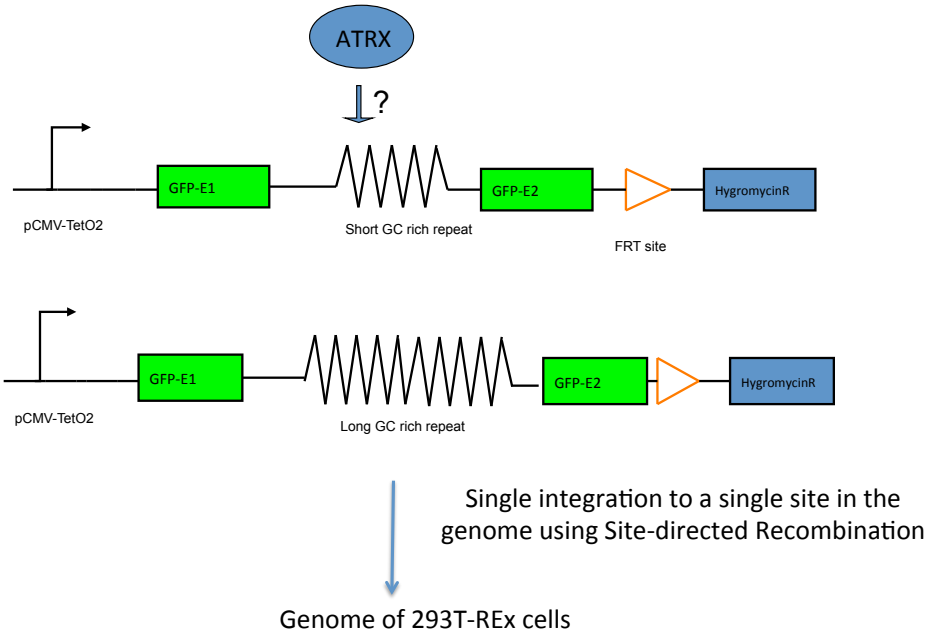

Figure 3.1 ***Experimental model****. GC-rich tandem repeats are inserted in an artificial intron of a reporter gene GFP with an inducible promoter pCMV-TetO2. This construct is then introduced into the genome of the 293T-REx cell line using Site-directed recombination. The repeats are varied in size, sequence and orientation regarding the G-rich strand.*

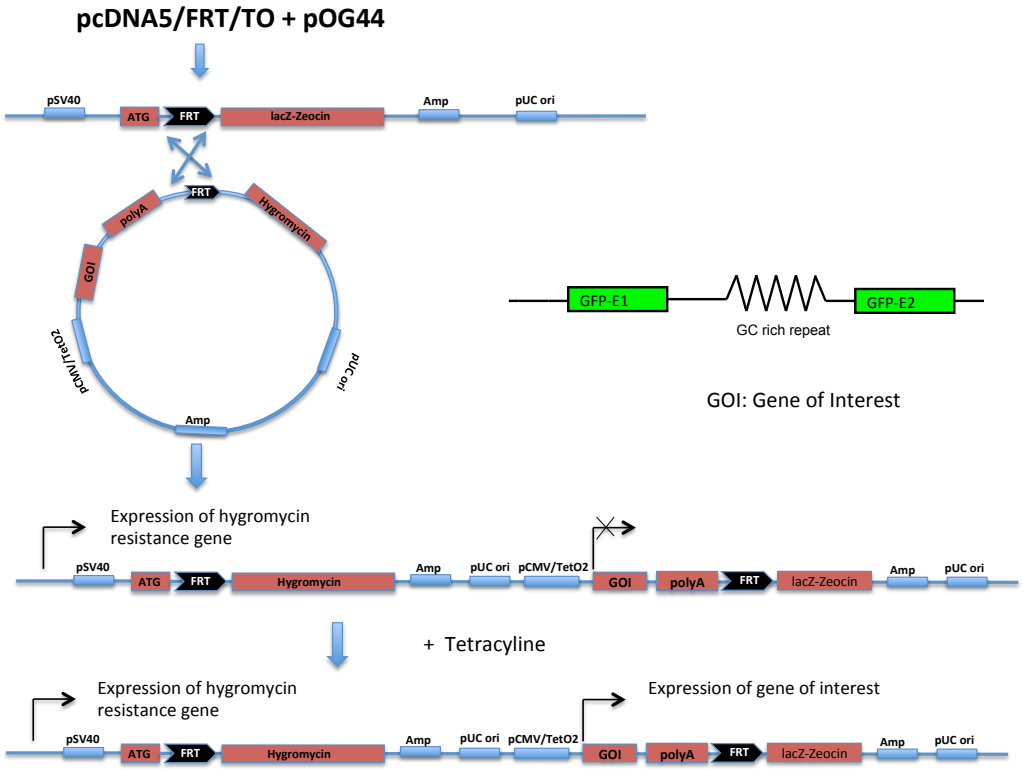

Figure 3.2 ***Schematic illustration of Flp-In 293T-REx stable expression system****. pcDNA5/FRT/TO containing gene of interest (GOI) is when co-transfected with pOG44, which expresses recombinase Flippase. When expressed Flippase facilitates intergration of the GOI into the genome of a host cell line that has an FRT site (between ATG codon and lacZ-Zeocin fusion gene) through site-directed recombination. In this project, the GOI is an artificial GFP gene with two exons (green) flanking an intron with a GC rich repeat. The result of a successful integration is that an entire construct of pcDNA5/FRT/TO is flanked by the ATG codon at the 5' end and lacZ-Zeocin without an ATG codon at the 3' end. Addition of tetracycline induces the expression of the GOI.*

Thirdly, the tetracycline (Tet) inducible promoter in this system allows investigation on the impact of transcription on ATRX targeting since the expression of the ectopic gene can be reversibly turned on or off in the presence or absence of the inducer, the

antibiotic tetracycline or its derivative doxycycline (which has the same mechanism of action as tetracycline), in the culture. The tetracycline inducible system, first developed by Gossen and Bujard in 1992 (see (Gossen and Bujard 1992)), is the most common inducible expression system used for studies in eukaryotes. It is comprised of two components, the Tet repressor and the Tet operator sequence. In nature, the *E. coli* Tet repressor (TetR) protein negatively regulates the genes of the tetracycline-resistance operon (TetO) on the Tn10 transposon (Hillen and Berens 1994). In the absence of tetracycline, every TetR homodimer binds to each TetO sequence with extremely high affinity, with a binding constant K_B of $2x10^{11}$ M^{-1}, as measured under physiological conditions (Hillen and Berens 1994). Binding of the homodimer represses the promoter bearing TetO.

In my experimental constructs, two TetO sequences embedded in the promoter of the reporter gene GFP (Figure 3.3) serve as the binding sites of 4 TetR molecules, intensifying the repression and reducing the basal expression of the gene. Upon addition, tetracycline binds to the TetR homodimers at a 1:1 ratio with an association constant of $3x10^{9}$ M^{-1} , causing a conformation change in the TetR that disassociates it from the TetO sequences (Hillen and Berens 1994), leading to gene derepression. Furthermore, the tetracycline inducible system has a number of advantages over other conditional expression systems based on Cre, FRT recombinases or ER (Estrogen Receptor). Activation of gene expression in the Tet system is reversible whereas it is irreversible in the Cre and FRT systems. Since the TetO sequence is naturally absent from mammalian cells, pleiotropy effect is believed to be minimal compared to the ER system.

In the first half of this chapter, I will describe the generation of constructs containing the GC-rich tandem repeat of the ψζ gene in the globin cluster (Proudfoot, Gil et al. 1982)(so called ψζVNTR). It will be inserted into the intron of the reporter GFP gene, which mimics its physiological genomic location (which is the second intron of the ψζ gene). To address the question as to how the length of the repeat influences ATRX recruitment, several ψζVNTR fragments with different sizes are cloned into the constructs.

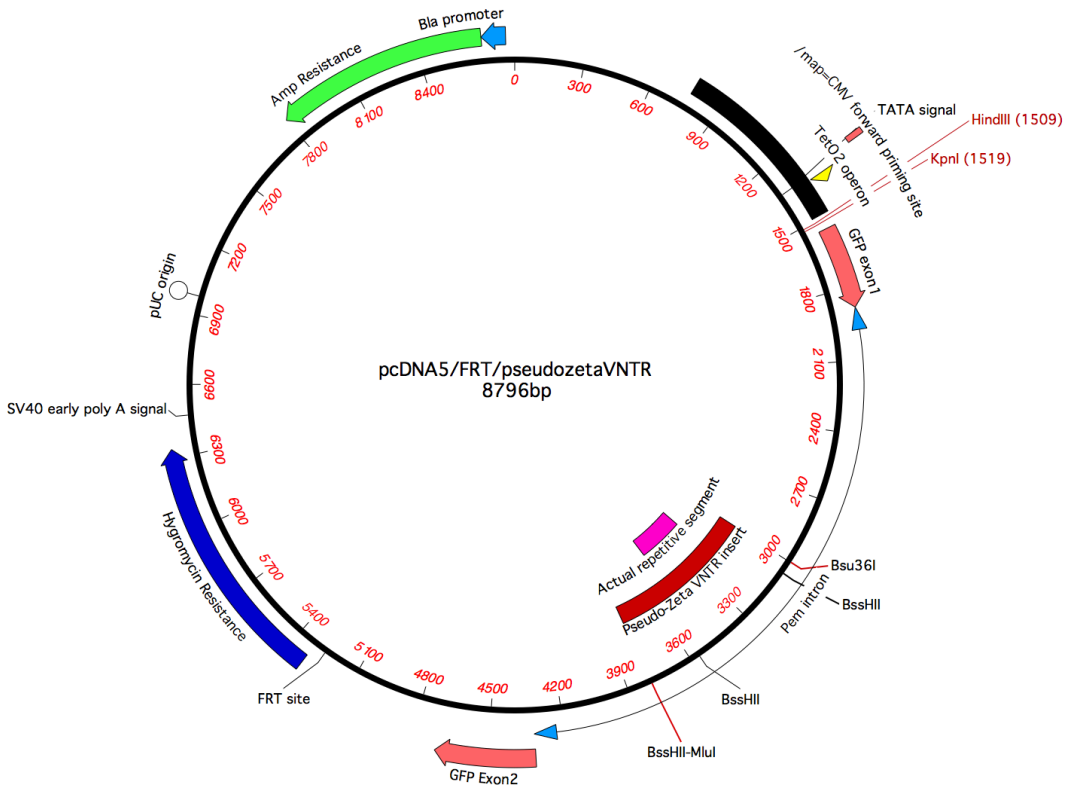

Figure 3.3 ***pcDNA5/FRT/ψζVNTR vector map****. ψζVNTR insert including non-repetitive sequences flanking the actual repetitive segment is engineered into the vector between the Bsu36I and the BssHII-MluI site.*

In order to answer the question as to whether it is the primary sequence or the secondary structure of GC-rich tandem repeats that determine ATRX targeting, it is

necessary to investigate ATRX binding in another GC-rich tandem repeat that has a primary sequence different from the ψζVNTR but shares the same properties such as the G-richness, GC skewness and G clustering. As discussed in Chapter 1, the human telomeric repeat, (TTAGGG)n, is a well-characterised target of ATRX (Law, Lower et al. 2010). Like ψζVNTR, it is also high in GC content and has G clusters. It also presents a positive GC skew, meaning there is a richness of guanine over cytosine, on one strand. Both types of repeat share features to form the same secondary structures such as G-quadruplexes and R-loops. This rationale leads to the second half of this chapter where I will describe the generation of stable expression 293T-REx cell lines containing ectopic human telomeric repeats.

In both types of constructs, the G-rich strand will be oriented in two ways with regard to the direction of transcription: in its physiological orientation on the non-template strand and in the reverse orientation on the template strand. This will provide materials for studying how the direction of transcription of the G-rich strand affects ATRX targeting.

3.2 RESULTS

3.2.1 Pseudo-zeta Variable Number Tandem Repeat (ψζVNTR) clones

3.2.1.1 Physiological orientation ψζVNTR clones

Since the ψζVNTR in the human alpha-globin locus was demonstrated as a target of ATRX, it was chosen to create the experimental model. To generate constructs containing repeats of different sizes, ψζVNTR from genomic DNA from 16 individuals was analysed by PCR using an amplicon that includes non-repetitive sequences flanking the repeat (Figure 3.4A).

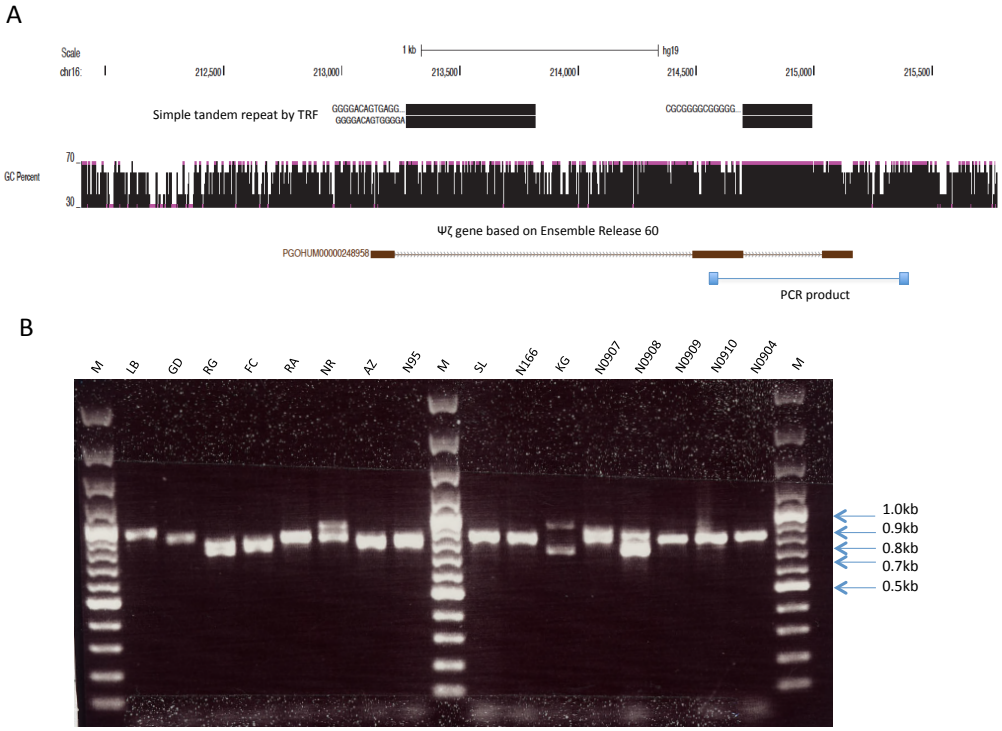

Figure 3.4 ***PCR amplification of the ψζVNTR*** *(A) A snap shot from genome browser shows the location of ψζ gene. The GC rich ψζVNTR of interest is situated in and spans the entire second intron of the gene. In silico PCR products are shown with the blue boxes at the ends representing the primers used for amplifying the repeat fragment. (B) ψζVNTR in 16 healthy individuals. Lane M: 2-log DNA marker (NEB N3200). Other lanes are marked with individual's name initials.*

As can be seen from Figure 3.4B, there are two genomic patterns of ψζVNTR; one is homozygous represented by a single band and the other is heterozygous represented by double bands indicating two copies of different sizes. The PCR product length ranges from 0.7kb to 1kb with the sequence as shown on Figure 3.5. Excluding the non-repetitive regions, it is estimated that the shortest repeat is about 140bp (derived from the second band from the top of lane N0908) and the longest is about 490bp (derived from the single band of lane LB). It is worth noting that the previously

documented size range of ψζVNTR in an ATR-X syndrome patient is from 330bp to 470bp (Law, Lower et al. 2010).

CCTGAGGCAAGACCTACTTCCCGCACTTCGACCTGCACCCGGGGTCCGCGCAGTTGCGCGCG
CACGGCTCCAAGGTGGTGGCCGCCGTGGGCGACGCGGTGAAGAGCATCGACGACATCGGC
GGCGCCCTGTCCAAGCTGAGCGAGCTGCACGCCTACATCCTGCGCGTGGACCCGGTCAACTT
CAAGGTGCGCGGGGCGCGGTGCGGGCGGGGCGGGGCGGGGCCGCGGGGCGGGCGGGGC
CGCGGGGCGGGGTCGCGGGGCGGGGCGGGGTGGGGTCGCGGGGCGGGGCGGGGTCGCG
GGGCGGGGCGGGGCGGGGCGGGGCGGGCGGGGCGGCCGGGGCCCGGCGGGGCGGGGC
GGGGCGGGGAGGGGCTGGGCGGGGCGGGGCGCGGGGCGGGGCGGGCCGGGCCGGGGC
GGGGTCGCGGGGCGGGGTCGCGGGGCGGGGCGCGGGGCGGGGCGGGGCGGGGTGGGG
TCGCGGGGCGGGGCCCGGGCTAGGCCCCGCCCCCGCACTGAGCCGCCCCCGCCCCCAGCTC
CTGTCCCACTGCCTGCTGGTCACCCTGGCCGCGCGCTTCCCCGCCGACTTCACGGCCGAGGC
CCACGCCGCCTGGGCCAAGTTCCTATCGGTCGTATCCTCTGTCCTGACCGAGAAGTACCGCT
GAGCGCCGCCTCCGGGACCCCCAGGACAGGCTGCGGCCCCTCCCCTGCCCTTCACCCTCCCA
CAGTTCCTGCCCTGACTCCAATAAATGGATGAGGACGGAGCGATCTGGGCTCTGTGTTCTCA
GTATTGGAGGGAAGGAACGCGT

Figure 3.5 ***ψζVNTR insert used in the model system****. The sequence of the PCR product including the ψζVNTR highlighted in blue, restriction site Bsu36I at the 5' end and MluI at the 3' end in red, which are used for cloning into pcDNA5/FRT/GFP vector.*

The products from two individuals LB and N0908 (designated N from now on) were used to clone into the plasmid pcDNA5FRT/TO/GFP (Figure 3.6A). The construction of this vector is described in Chapter 2 Materials and Methods. The VNTR was engineered into an exogenous intron, which was derived from the first intron of the rat Pem gene, located between two exons of the GFP gene. To reduce the likelihood of rearrangement, DH5α *E.coli* strain (Invitrogen) was used for the cloning. The insertion of the ψζVNTR was verified by restriction enzyme (BssHII) digestion on the isolated plasmids (Figure 3.6B and C). The homozygous LB fragments gave rise to clones with the same size of ψζVNTR (Figure 3.6B). The heterozygous N fragments produced clones with three different lengths of the VNTR, one of which might result from a deletion that occurred during cloning (Figure 3.6C). After excluding the non-

repetitive regions between the BssHII sites to the ends of the actual repeat, it is estimated that clones LB1 to LB8 contain an approximate 490bp repeat, clones N1, N4, N5, N7 and N8 contain an approximate 240bp repeat, clone N2 140bp, and clone N3 390bp. Following cloning, to confirm that the inserts were indeed ψζVNTR, clones LB8, N1, N2 and N3 were selected for sequencing using a forward primer (Upstream-Int F) and reverse primer (Downstream-Int R).

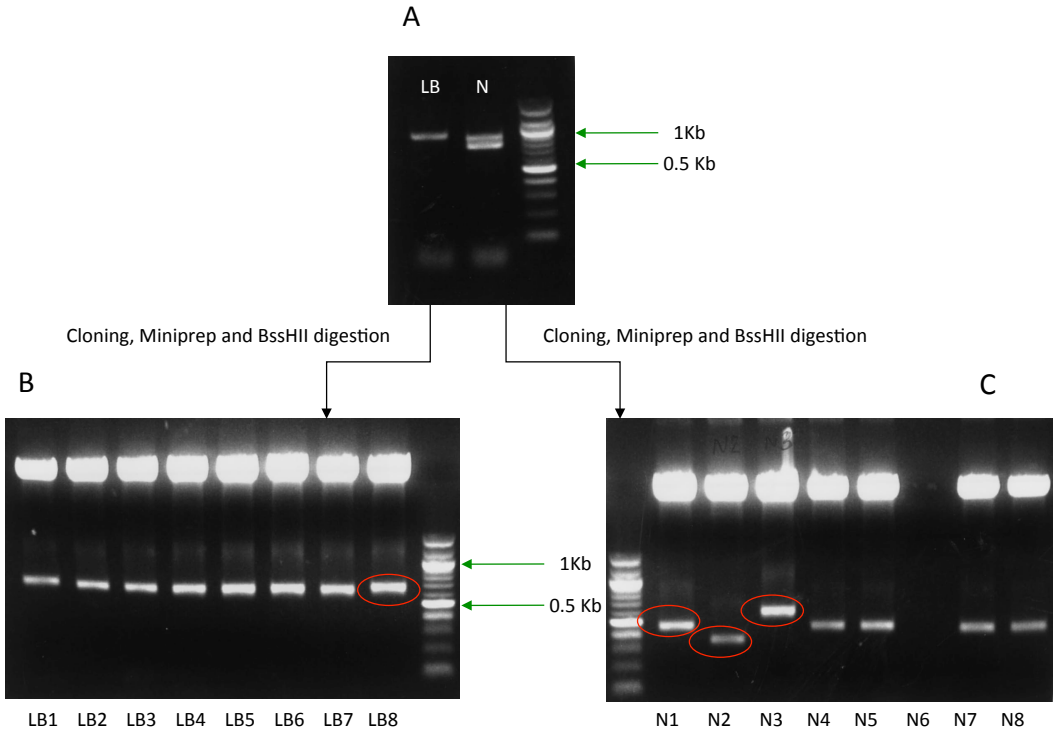

Figure 3.6 ***Subcloning of the ψζVNTR****. (A) PCR products of ψζVNTR from genomic DNA of two healthy individuals (LB and N). LB is homozygous in the ψζVNTR (single band at around 1kb) whereas N is heterozygous with two copies of ψζVNTR of different sizes (approximately 0.9 and 0.8kb). (B) and (C) BssHII digest to check the insertion of LB and N PCR products into pcDNA5/FRT/GFP. As expected, cloning of LB product results in clones (LB1 to LB8) with ψζVNTR inserts of the same size (at around 0.7kb). Cloning of N products, however, gives three kinds of clones; one with a ψζVNTR insert of about 650bp (N3), another one with 450bp (N1, N4, N5, N7 and N8) and the other with 350bp (N2).*

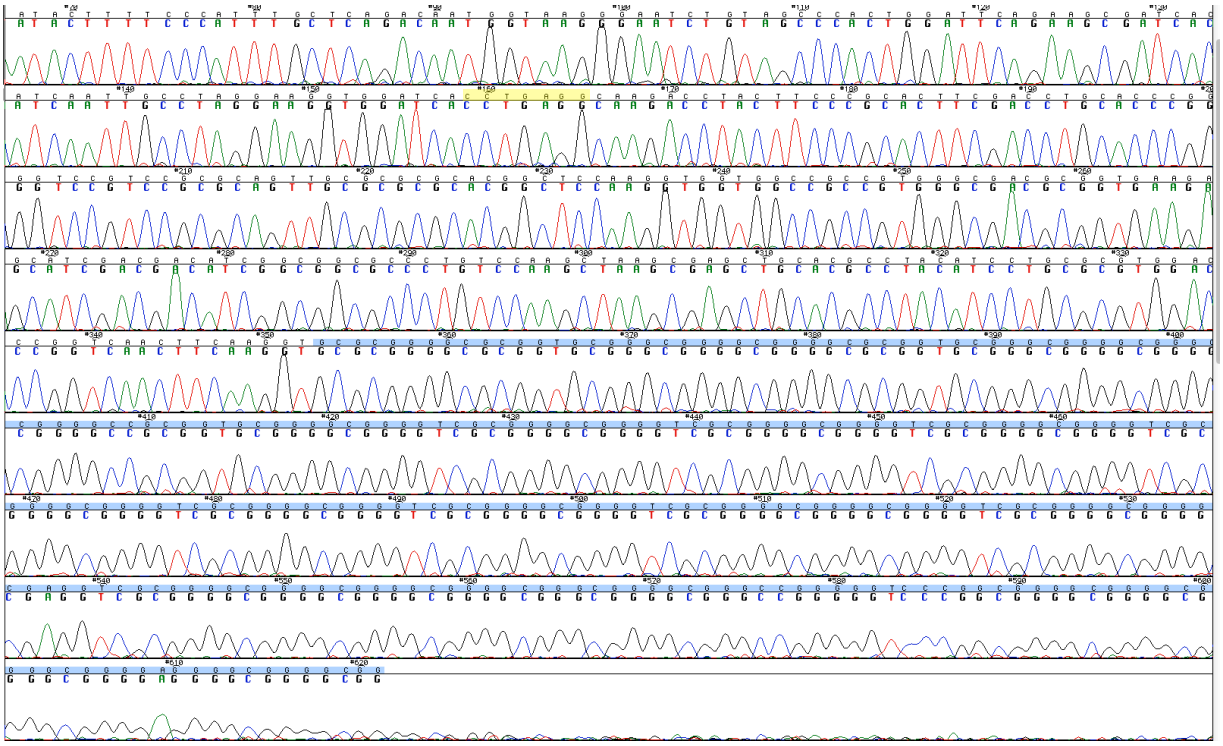

Figure 3.7 ***The sequencing result of a plasmid from one of the bacterial clones successfully harboring ψζVNTR*** *(highlighted in blue). This is the reading from forward primer Upstream IntF, which shows the restriction site Bsu36I at 5' end of the repeat highlighted in yellow.*

Sequencing this GC rich tandem repeat was extremely difficult. After many rounds of optimization, addition of DMSO into the sequencing PCR reaction mix was found to significantly improve the results. However, due to long stretches of G and C bases, I could only read about 200-250bp into the repeat at best, from the forward primer (Figure 3.7). With the reverse primer, the reads were even shorter. Nevertheless, these results were sufficient to confirm that the inserts were indeed the ψζVNTR.

To create a control construct carrying no repeats, clone N2 was digested with BssHII. The digest was run on an agarose gel and the vector backbone was recovered. This

linearized vector containing BssHII overhangs at both ends was re-circularised by T4 ligase to give rise to a construct without the repeat, designated N2S.

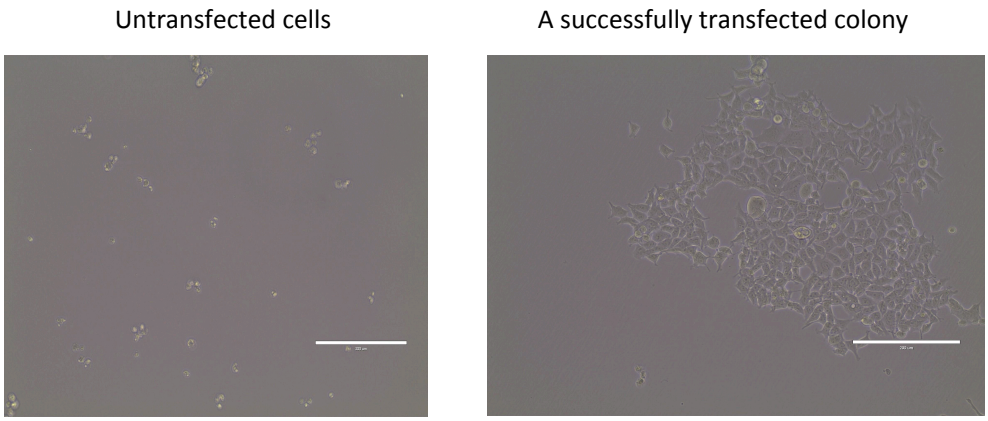

Figure 3.8 ***Selection of stable expression 293T-REx in medium containing hygromycin B****. Untransfected (host) cells die from hygromycin B treatment whereas cells successfully harboring desired constructs form colonies.*

Introduction of the constructs into the cell line was carried out as described in Materials and Methods. Incorporation of a construct into the correct site in the host cell line results in three interdependent events; The fusion gene LacZ-Zeocin is disrupted; the hygromycin resistance gene, which originally lacks a promoter and a start codon on the construct, is now activated because it is placed in-frame adjacent to the start codon and pSV40 promoter, which was previously part of the fusion gene LacZ-Zeocin. The LacZ-Zeocin fusion gene is hence separated from its start codon and promoter, and becomes inactive. Clones successfully harboring the constructs were identified and selected by a combination of three assays, hygromycin resistance (Figure 3.8), zeocin sensitivity (Figure 3.9) and lack of beta-galactosidase activity (Figure 3.10).

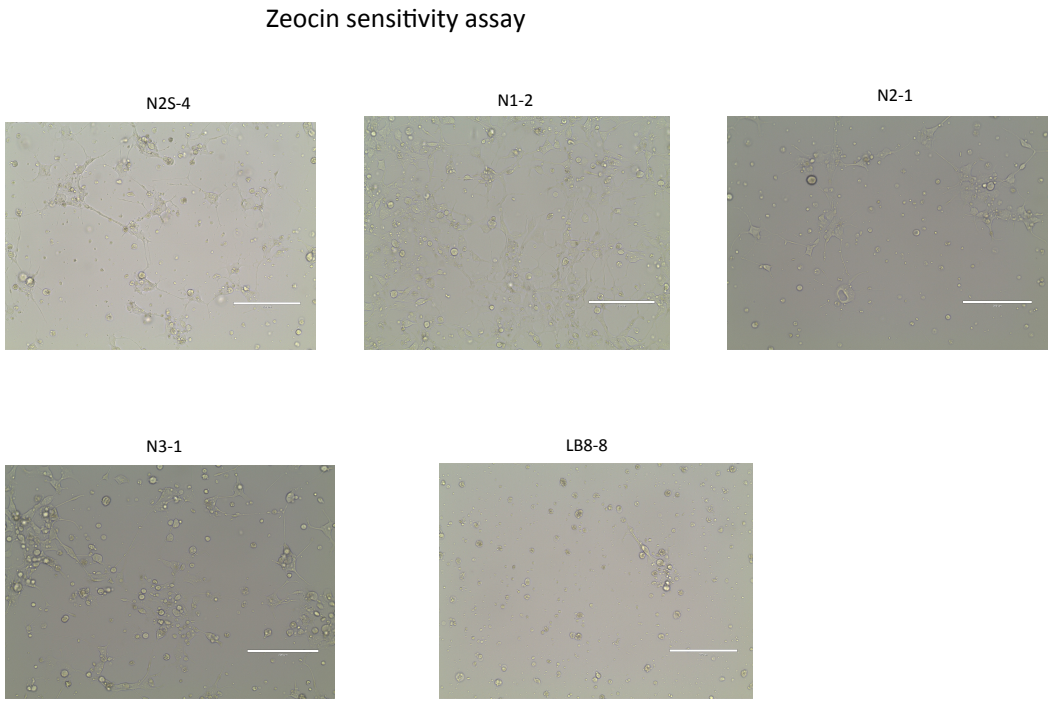

Figure 3.9 ***Zeocin sensitivity assay****. Untransfected host cells Flp-In 293T-REx grow as normal in medium containing Zeocin because their Zeocin resistance gene, which is fused with LacZ gen, is intact and active. Successfully tranfected 293T-REx with the ψζVNTR constructs die since their fusion gene is disrupted and inactive. The images show the successfully transfected clones.*

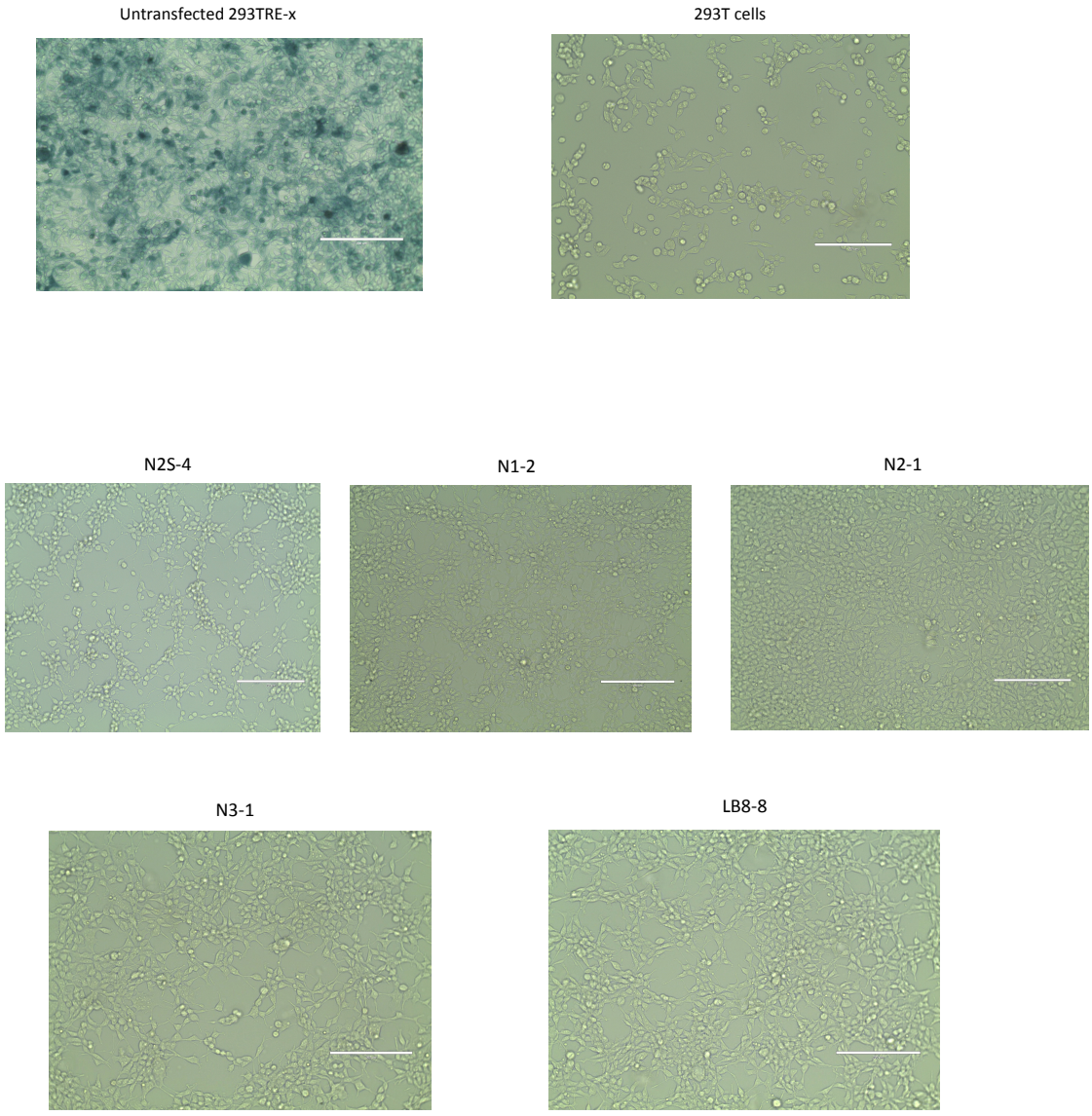

Figure 3.10 ***Beta-Gal staining selects positive clones that have successfully incorporated the ectopic constructs****. Untransfected 293T-REx host cells stain blue because LacZ gene is still intact and active. 293T cells, however, do not stain since they do not have an intact LacZ gene. Images of representative clones (N2S-4, N1-2, N2-1, N3-1 and LB8-8) harboring constructs containing ψζVNTR of different sizes.*

To confirm that the integration occurs as a single event, Southern Blot analysis was carried out for all the assay-selected clones. Genomic DNA was isolated, digested with a single cutter KpnI, and then probed with GFP exon 2 (Figure 3.11A). All of the analysed clones, except the host cell line 293T-REx, showed a single band on Southern Blot (Figure 3.11B and C), indicating a single integration of the cassettes.

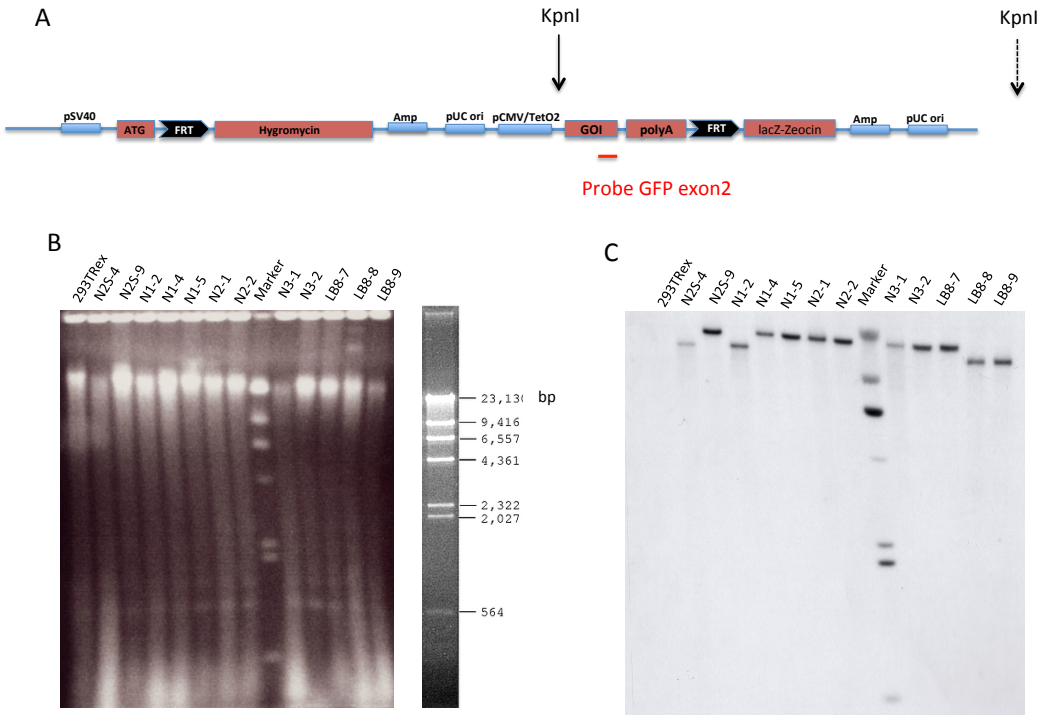

Figure 3.11 ***Verification of single integration of the ectopic ψζVNTR constructs by Southern Blot analysis****. (A) Diagram shows the genomic ectopic ψζVNTR cassette. A single internal KpnI site is located near GFP gene (GOI), indicated by a solid arrow and an external site is indicated by a dotted arrow. The red line indicates the position of the probe used for Southern Blot analysis. (B) The agarose gel image shows a KpnI digest of genomic DNA from the indicated clones. Pattern of the marker Lamda DNA HindIII digest (lane "Marker") is indicated as on the left of the gel. (C) Southern Blot assay of the KpnI digest in (B) using P32 labeled probe GFP exon2 (red line in A) shows single bands, suggesting single integration events, in the indicated clones except for the untransfected 293T-REx.*

To verify that the constructs were incorporated into the genome at the specific FRT site and not at a random location, a PCR analysis was performed using the primer pair pSV40 F/Hygromycin R (Figure 3.12). A correct incorporation gives rise to a product of 278bp in length whereas integration at a random site or failed integration will not result in any product or if any, of a undesired size. Figure 3.12 shows that all the selected clones have a correct integration.

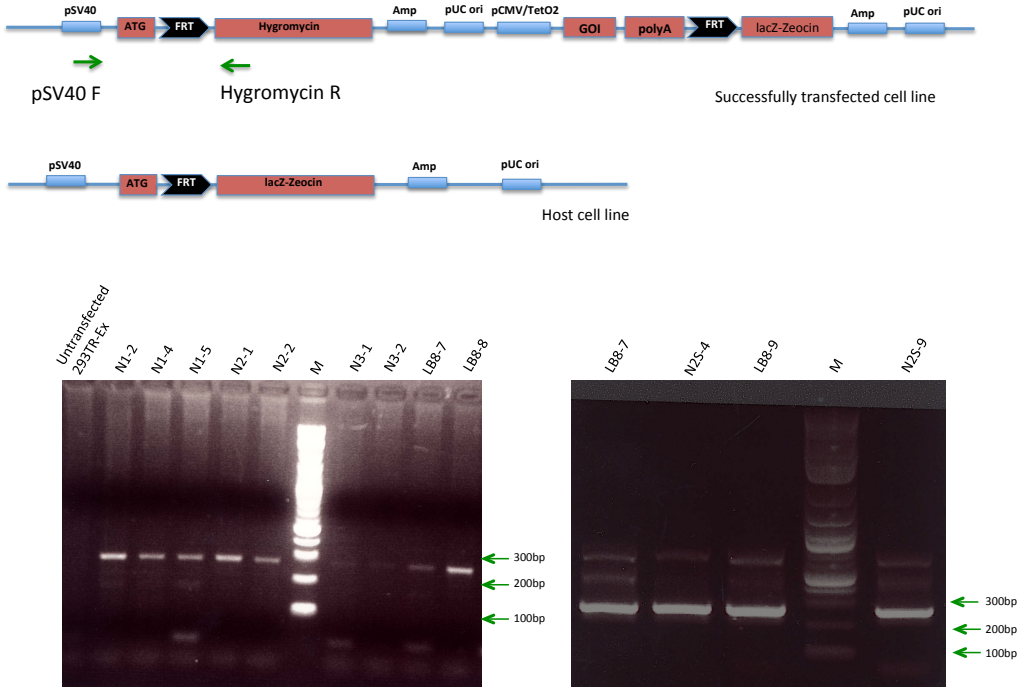

Figure 3.12 ***Confirmation of specific integration of the ectopic ψζVNTR constructs into the FRT site by PCR****. The top diagram shows the genomic ectopic ψζVNTR constructs with the green arrows indicating the primers used for checking specific integration at the FRT site. The bottom diagram shows the FRT site in the host (untransfected) cell line. The agarose gel images show the PCR products (major bands at about 270bp) in the indicated clones, suggesting that the integration occurred specifically at the desired FRT site.*

To further demonstrate the presence of the ectopic VNTRs in the clones, a PCR analysis of genomic DNA using the primer pair Upst IntF/GFP IntR specific for the ectopic ψζVNTR, but not the endogenous VNTR, was performed. Figure 3.13 shows the ectopic VNTR in clones of several constructs with different sizes of ψζVNTR as expected. Sequencing the PCR products confirmed that these were indeed ψζVNTR. At least two clones of each size of the ψζVNTR were chosen for investigating ATRX binding; no VNTR (no repetitive region but only the flanking region, N2S-4, N2S-5 and N2S-9), 140bp VNTR (N2-1 and N2-2), 240bp VNTR (N1-2, N1-4 and N1-5), 390bp VNTR (N3-1 and N3-2) and 490bp VNTR (LB8-7, LB8-8 and LB8-9).

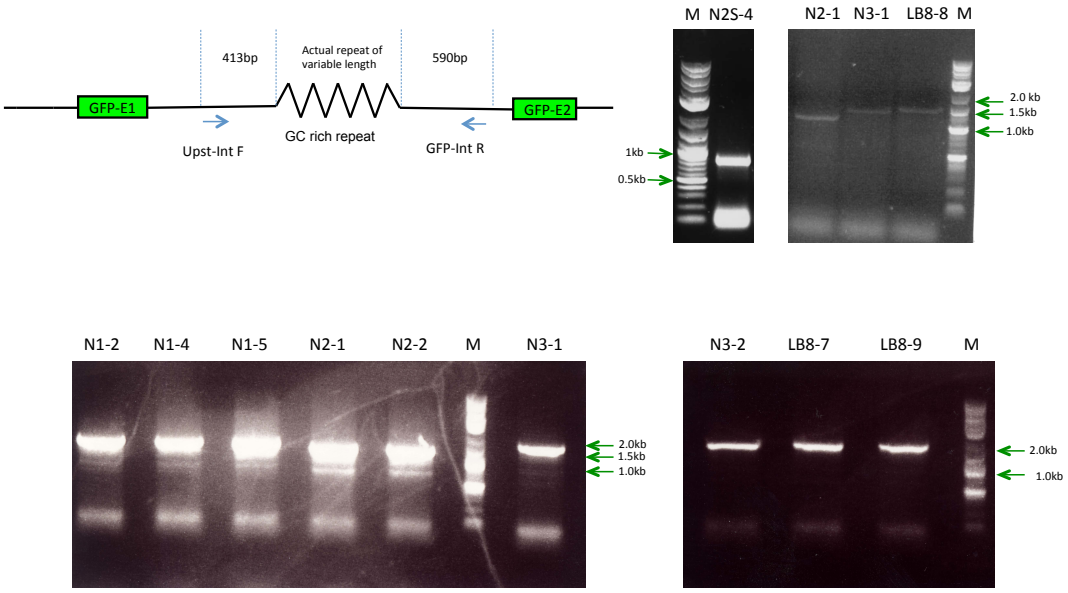

Figure 3.13 ***Confirmation of the presence of the ectopic ψζVNTR insert in the genome by PCR****. The diagram shows the positions of the primers used for PCR checking the ectopic ψζVNTR presence relative to the actual repeat. Amplification products from the genomic DNA of the indicated clones were run on 1%TAE gel. 2-log DNA ladder (NEB N3200) is used as a marker.*

3.2.1.2 Reverse orientation ψζVNTR clones

As can be seen from Figure 3.5, the G-richness of the ψζVNTR repeat is skewed on one strand, the non-template strand. It is called the physiological orientation of the repeat. To understand how the orientation of the G-richness affect ATRX recruitment, the repeat is inverted so that the non-template strand is now C-rich.

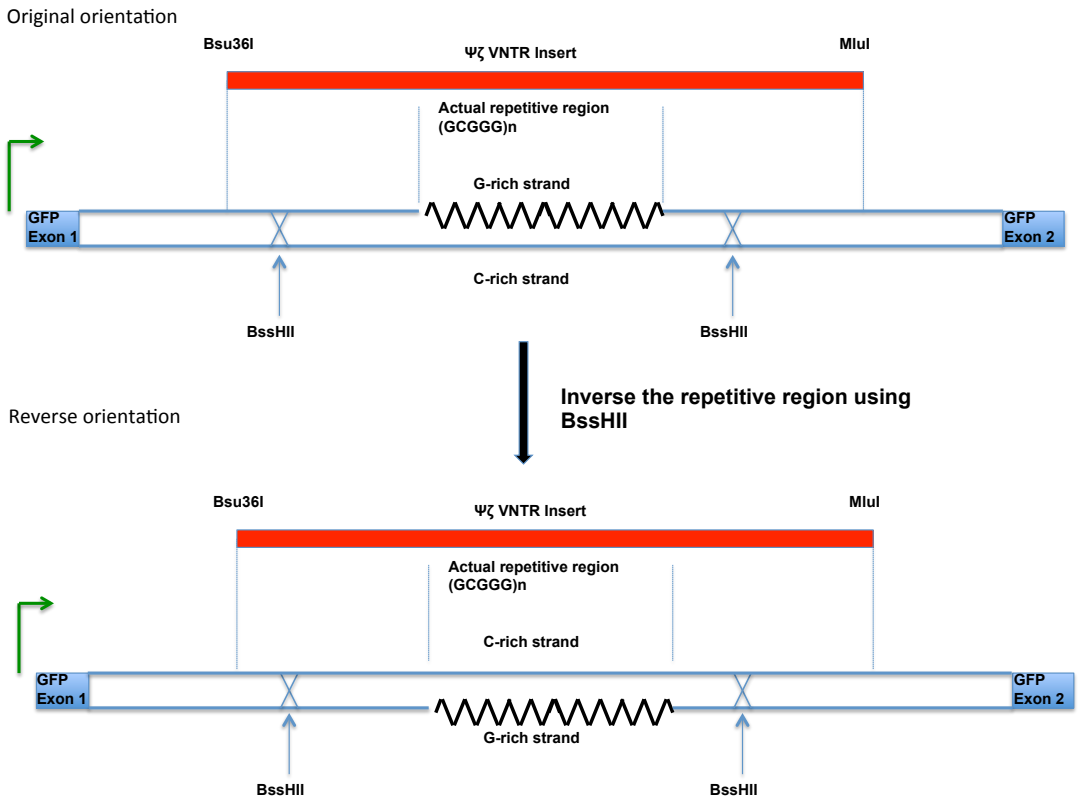

Figure 3.14 ***Scheme depicts of the ectopic ψζVNTR inversion strategy.***

Figure 3.14 describes the inversion strategy. The repeat fragment from clone N1 and clone LB8 was sliced out by BssHII digestion. The repeat fragments and the vector backbone were then separated and recovered from an agarose gel. Following dephosphorylation by Shrimp Alkaline Phosphatase, the vector backbone was re-

ligated to the repeat fragment and cloned into Stbl2. There is a 50% probability that the resulting colonies contain the repeat of the reverse orientation.

The only way to select these clones is to sequence them using the primer Downstream IntR. The sequence obtained from a successfully inverted construct with the reverse primer Downstream IntR should be identical to the one obtained from the original construct with the forward primer Upstream InF. One example of a successful reverse construct is shown on Figure 3.15.

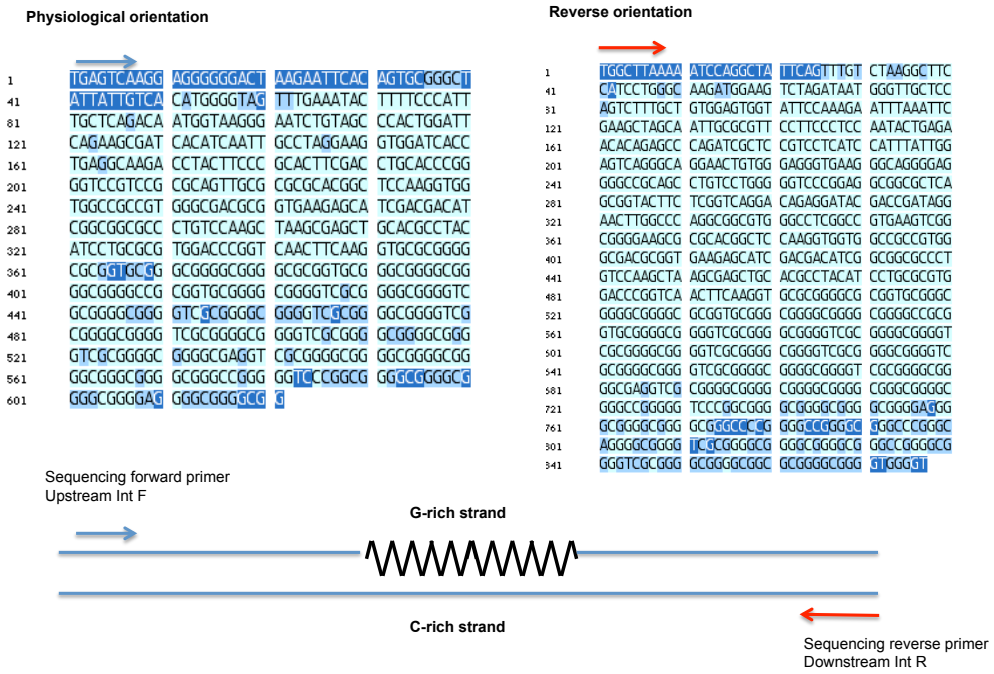

Figure 3.15 ***Comparison of sequencing results of two constructs carrying ψζVNTR of physiological orientation (left) and of reverse orientation (right).*** *The construct on the left (physiological orientation) was sequenced using the forward primer Upstream IntF, and the construct on the right (reverse orientation) was sequenced using the reverse primer Downstream IntR.*

Plasmid miniprep of a reverse short ψζVNTR construct (denoted N1R) and a reverse long ψζVNTR (denoted LB8R) was used to generate stable expression 293T-REx clones. Following hygromycin resistance and zeocin sensitivity assay, selected clones (clone 2, 3, 4 and 5 for N1R; clone 1, 2, 3, 4, 5, 6, 8 and 9 for LB8R) were screened by Southern Blot for single integration (Figure 3.16) and by PCR analysis of the FRT site for specific integration (Figure 3.17). AflII, instead of KpnI, was used to digest genomic DNA for Southern Blot analysis this time, simply because it was thought to give a more efficient digestion.

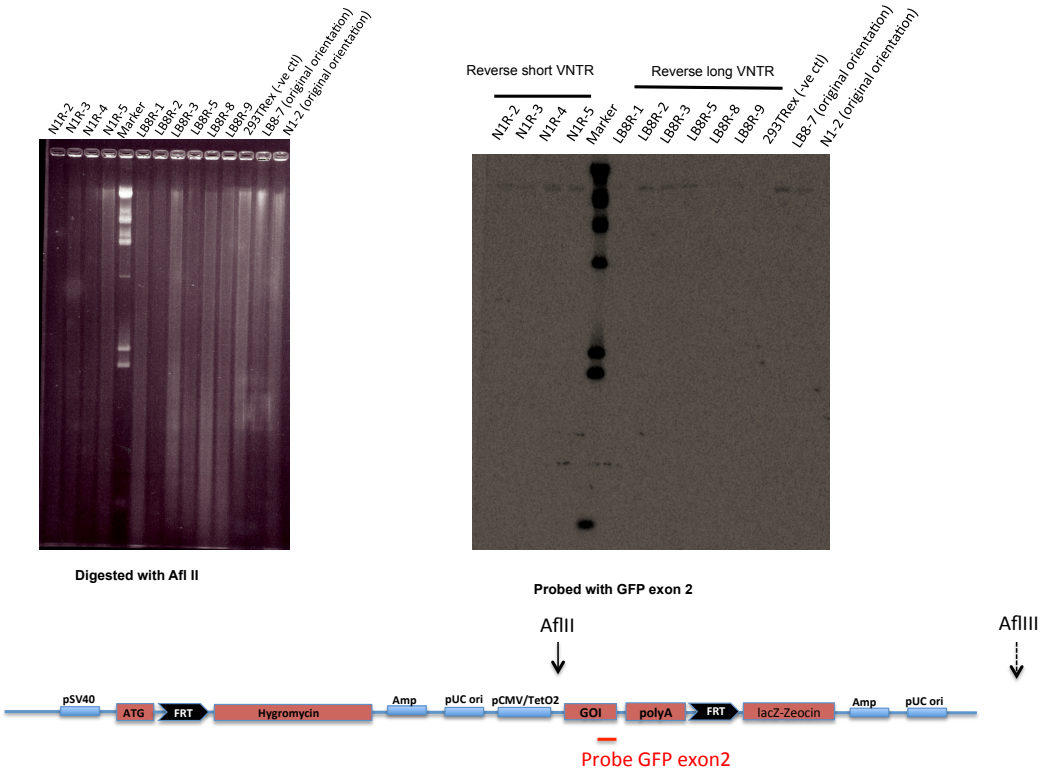

Figure 3.16 ***Southern Blot analysis that the integration of the reverse ψζVNTR in the indicated clones is a single event****. Genomic DNA of the indicated clones was digested with AflII, which cuts once in the ectopic construct fragment (solid arrow) and somewhere outside of the fragment (dotted arrow) and then probed with P32 labeled GFP exon 2. The marker is Lamda DNA HindIII digest (NEB).*

The Southern Blot results show a clear single band in all of the clones, except for LB8R-1, LB8R-8 and LB8R-9 showing weak signal, implicating single integration compared to the negative control (unstransfected 293T-REx) and the positive controls, which are clones containing ψζVNTR of the physiological orientation (LB8-7 and N1-2) (Figure 3.16). These integration events occurred at a specific FRT site as suggested by the results of the FRT site PCR analysis (Figure 3.17), except for clone LB8R-4 and LB8R-6.

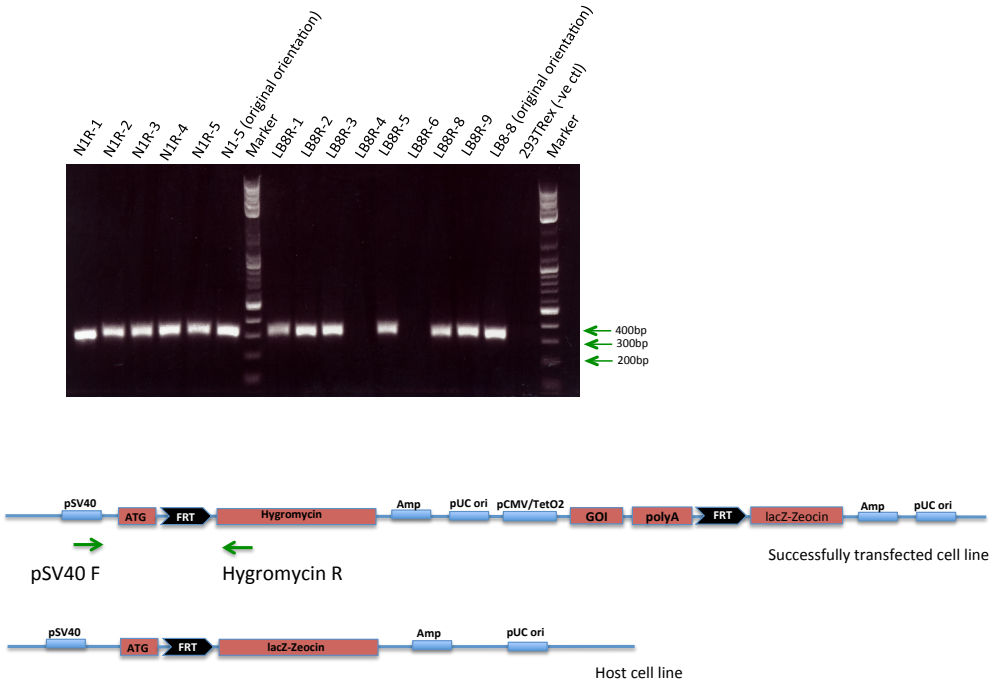

Figure 3.17 ***Confirmation of successful integration of ectopic reverse ψζVNTR constructs*** *at the specific FRT site by PCR analysis using the primers indicated by green arrows.*

All of the clones, except LB8R-6, showed the existence of the ectopic reverse ψζVNTR at the expected sizes (Figure 3.18) in the PCR analysis using the primer pair Upst-IntF and GFP-IntR, which was followed by sequencing using the same primers.

Clone N1R-4, N1R-5, LB8R-1, LB8R-2, LB8R-3 and LB8R-5 were chosen for ATRX binding analysis.

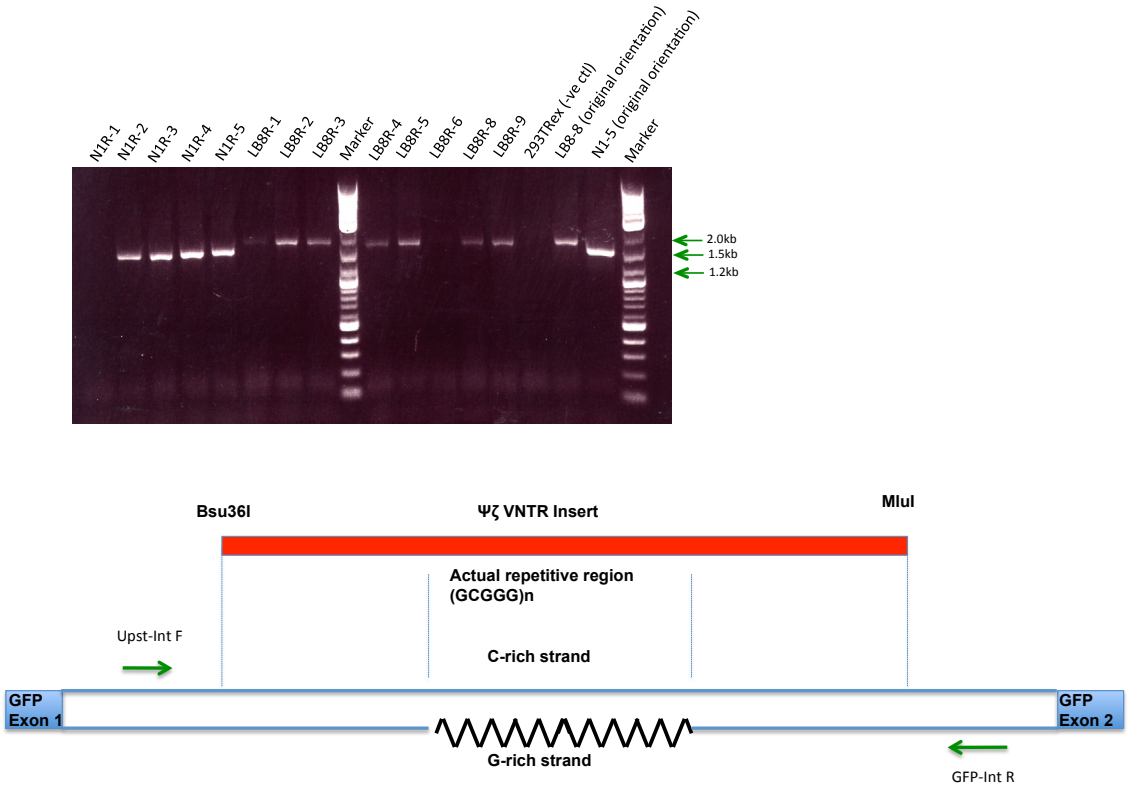

Figure 3.18 ***Confirmation of the presence of the reverse ectopic ψζVNTR*** *in the indicated clones using primer Upst-Int F and GFP-Int R. The marker used is 2-log DNA ladder.*

3.2.2 Telomeric repeat clones

3.2.2.1 Physiological orientation telomeric repeat clones

To create constructs containing telomere repeats (Figure 3.19), a Homologous Recombination based cloning method, called CloneEZ (GenScript), was employed (Figure 3.20C).

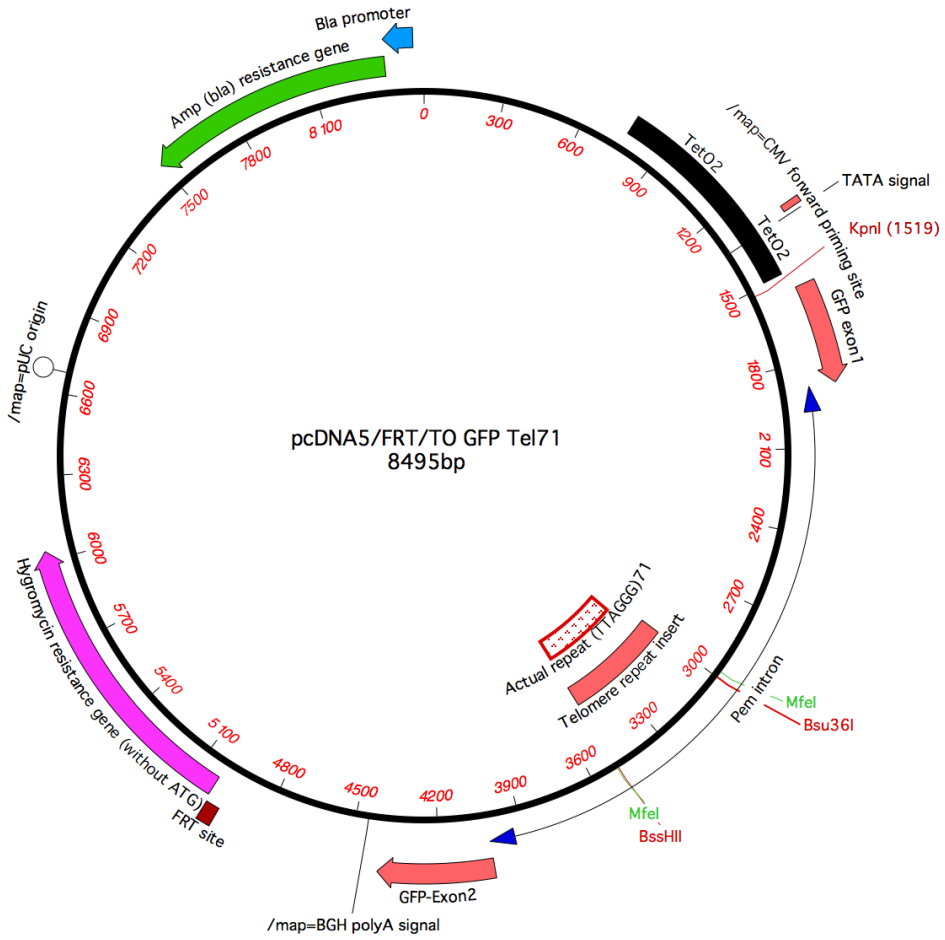

Figure 3.19 ***Map of the vector containing a telomeric sequence (of 71 repeat units) in the intron of the GFP gene pcDNA5/FRT/Tel71****, also called pcDNA5/FRT/(TTAGGG)71. The telomeric sequence is inserted into the vector between Bsu36I and BssHII site.*

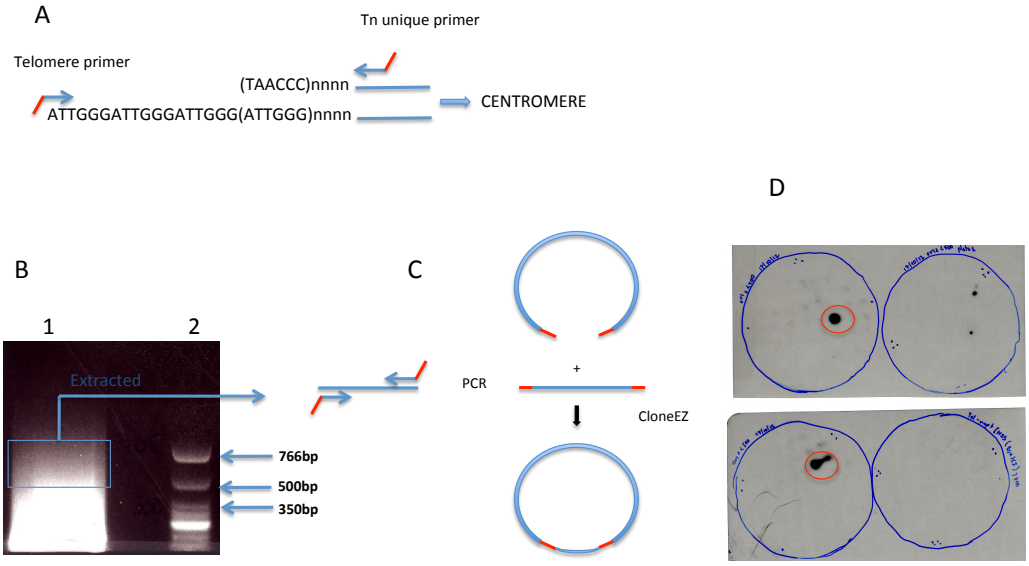

Figure 3.20 ***Cloning of human telomeric repeats using homologous recombination based-method (CloneEZ, GenScript)****. (A) Schematic illustration of designed primers with add-on sequences (red) that are homologous to the pcDNA5/FRT/GFP linearised vector and their position on the DNA template on a truncated chromosome with a fusion telomere. (B) Lane 1: PCR product of telomeric repeats using the above primers, lane 2: Low Molecular Weight Ladder (NEB, N3233). (C) Schematic illustration of CloneEZ method. The PCR product is integrated into the linearised vector by homologous recombination using the homology arms (red). (D) Positive clones harbouring the insert (the PCR product of telomeric sequence) detected by colony lift assay.*

First, vector pcDNA5/FRT/GFP was linearised by Bsu36I and BssHII. Telomeric repeat insert fragments were obtained by PCR amplification using primers with a 15 base add-on sequence homologous to either side of the restriction site (Figure 3.20A). Genomic DNA from the patient bearing a truncated chromosome with a fusion telomere (kindly provided by Christian Babbs) was used as the template for the PCR. The products which are of different lengths form a smear on the agarose gel (Figure

3.20B) since the Forward primer (Telomere primer, containing the repeat sequence and an add-on sequence annealed unspecifically along the repeat fragment.

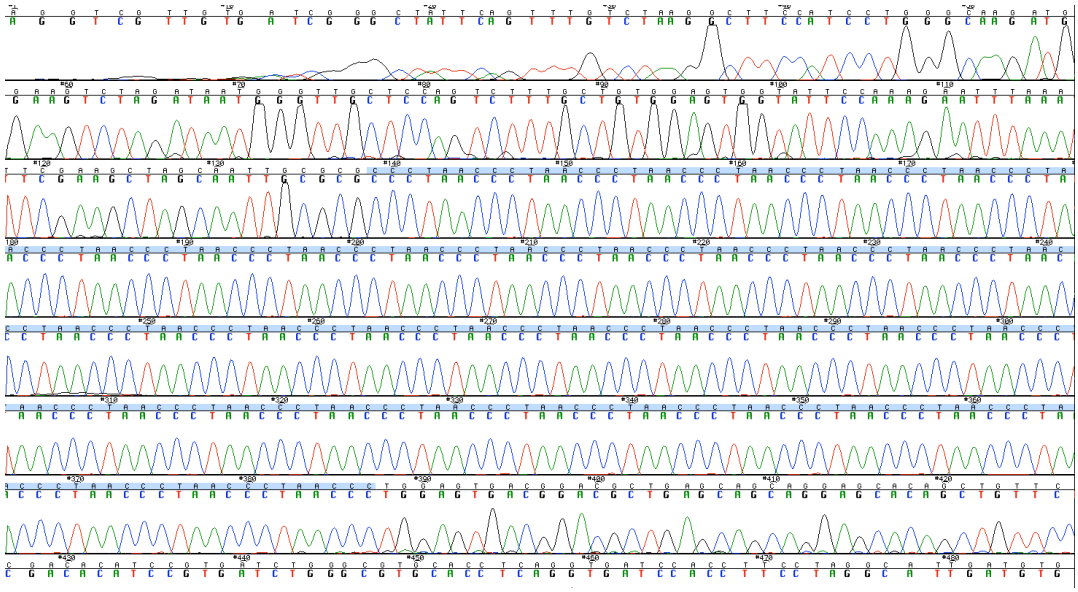

Figure 3.21 ***Sequencing result of a plasmid containing a telomere repeat of 42 units (TTAGGG)42****.*

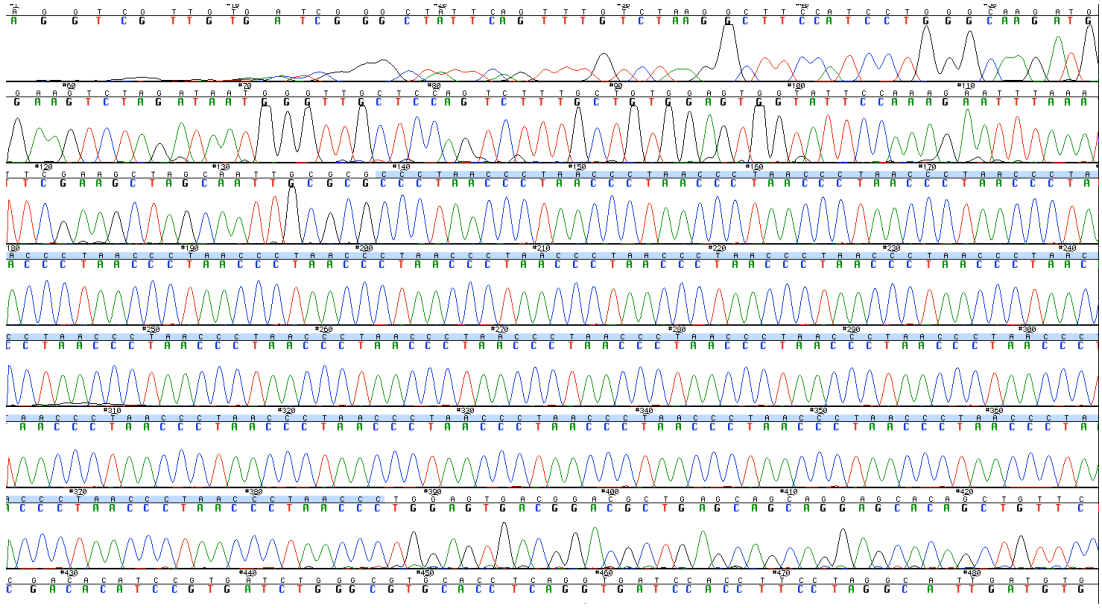

Figure 3.22 ***Sequencing result of a plasmid containing a telomere repeat of 71 units (TTAGGG)71.***

The products ranging in size from 500bp to 700bp were isolated from the gel and were then recombined with the linearised plasmid, which was followed by transformation into Stbl2 competent cells. To screen for positive colonies harboring recombined plasmids, colony lift assay was performed (See Chapter 2 Materials and Methods). Two colonies were identified and were subjected to sequencing (Figure 3.20D). Colony 1 was confirmed to have 42 units of telomeric repeat (TTAGGG)42 (also called Tel42), and colony 2 71 units (TTAGGG)71 (also called Tel71) (Figure 3.21 and Figure 3.22). Plasmids were isolated from these colonies and were then introduced into 293T-Rex.

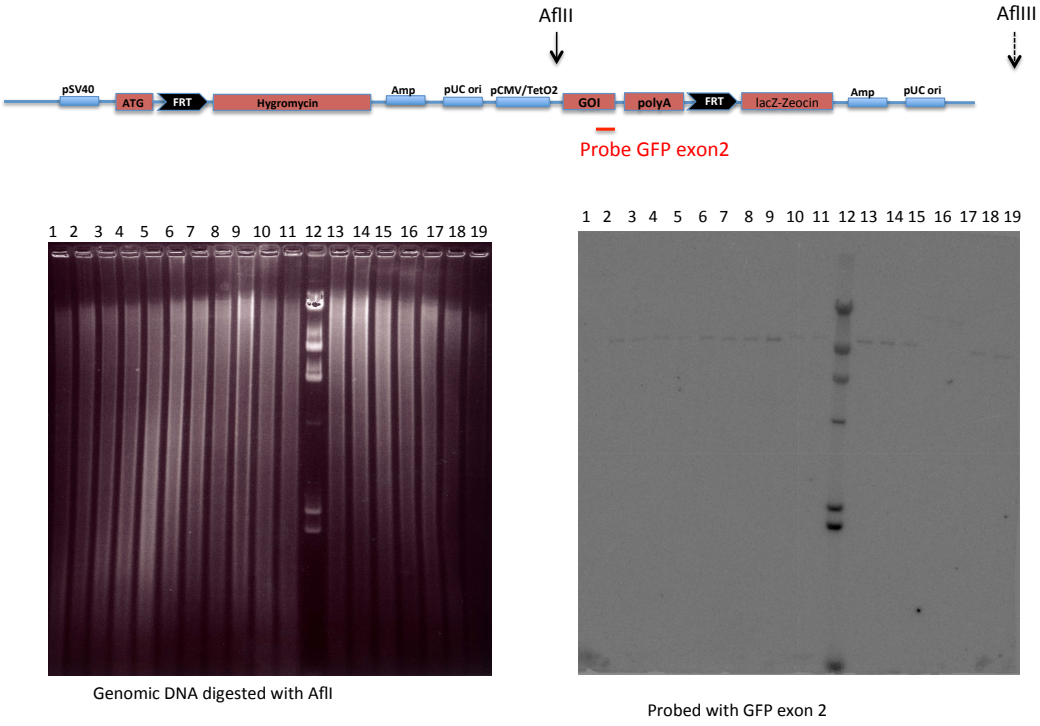

Figure 3.23 ***Southern Blot verifies single integration of telomere repeat cassettes in corresponding clones****. Lane 1 untransfected 293T-REx, lane 2 LB8-7, lane 3-6 GAA clone 2,4, 6 and 7 respectively, lane 7-11 Tel42 clone 7, 6, 5, 3 and 1 respectively, lane 12 marker, lane 13-19 Tel71 clone 7, 6, 5, 4, 3, 2 and 1 respectively.*

Hygromycin resistant, zeocin sensitive clones were screened by Southern Blot and FRT PCR for single and specific integration (Figure 3.23 and Figure 3.24A). Three clones of each construct (clone 5, 6 and 7 of Tel42; clone 5, 6, and 7 of Tel71) were verified by PCR (Figure 3.24B) and sequencing to have an ectopic telomeric repeat in their genome.

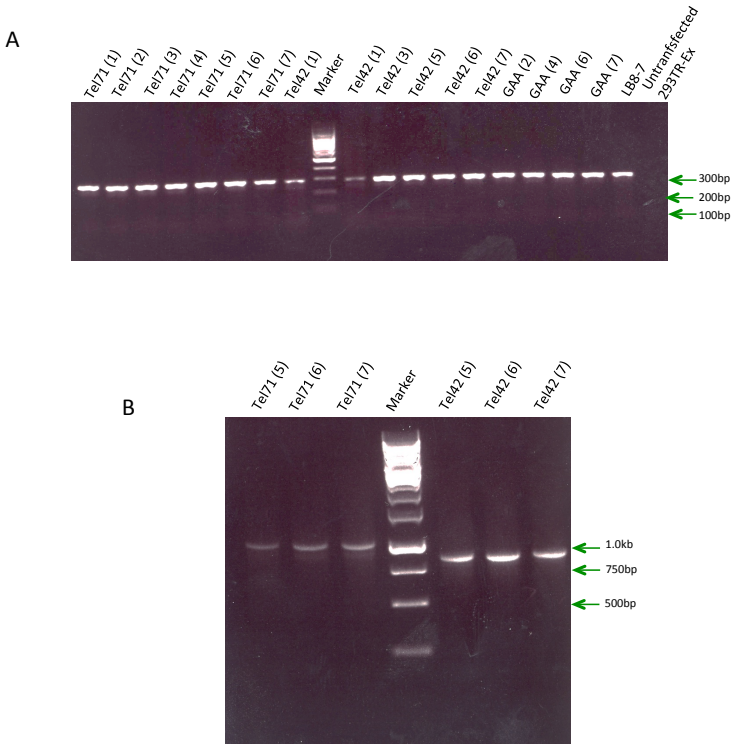

Figure 3.24 ***Verification of telomere repeat containing clones****. Confirmation of successful integration of ectopic telomeric constructs at the specific FRT site by PCR using the primer pair pSV40F/Hygromycin R. (B) Confirmation of the presence of ectopic telomere repeat in the genome of indicated clones by PCR using the primer pair UpstreamIntF/DownstreamIntR.*

3.2.2.2 Reverse orientation telomeric repeat clones

To create reverse orientation telomeric repeat clones, my strategy was to invert the repeat fragment in the Tel42 and Tel71 constructs and then introduce these reverse

orientation constructs into 293T-Rex. The inversion was carried out using two MfeI restriction sites flanking the repeat (Figure 3.25). Digestion with MfeI and religation of the fragment into the vector backbone led to two possibilities; the fragment restored or reversed its original orientation.

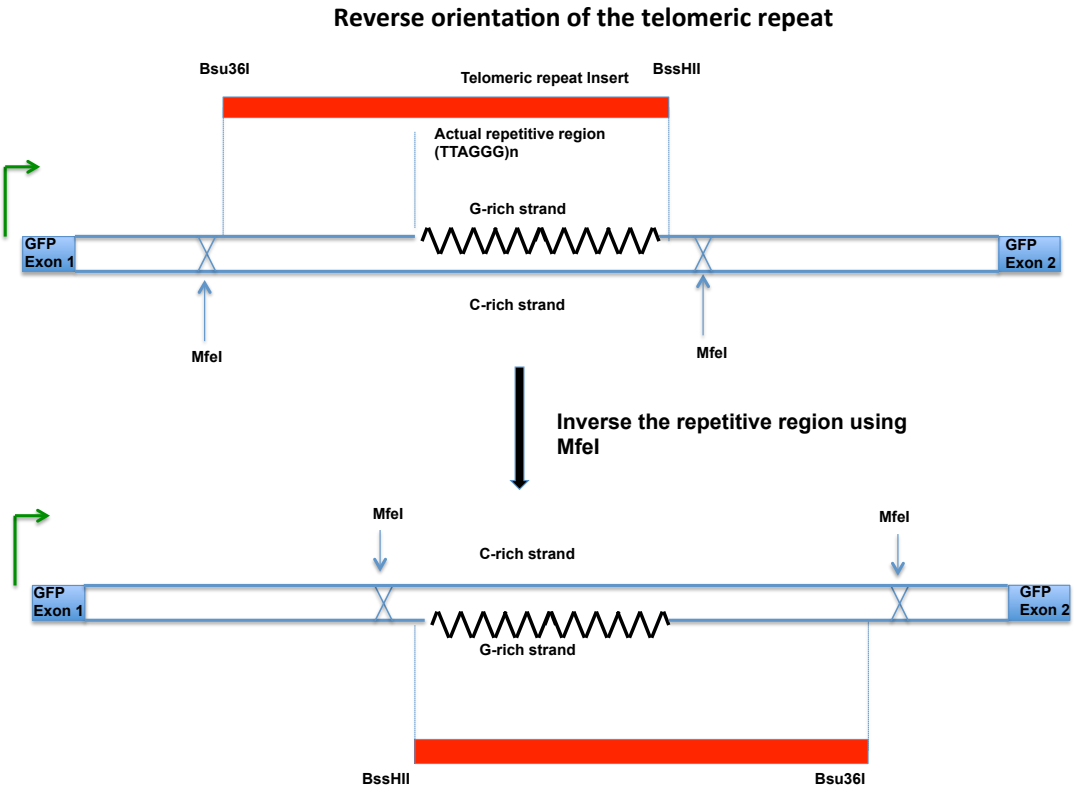

Figure 3.25 ***Inversion of the telomere repeat using MfeI restriction sites flanking the repeat insert****. The telomere repeat insert defined by the fragment between Bsu36I and BssHII (red block) restriction sites has G-richness on the non-template strand in its original (physiological) orientation. Excision of the fragment and religation back into the vector can reverse the orientation of the fragment so that G-richness is now on the template strand.*

To identify colonies bearing reverse orientation, isolated plasmids were double digested with Bsu36I and SwaI. Successfully inverted constructs (designated Rtel71

and Rtel42) will result in a fragment of 12bp, which is too small to be visualized on a gel, whereas original Tel71 and Tel42 constructs digest will give rise to 500bp and 400bp fragment, respectively (Figure 3.26).

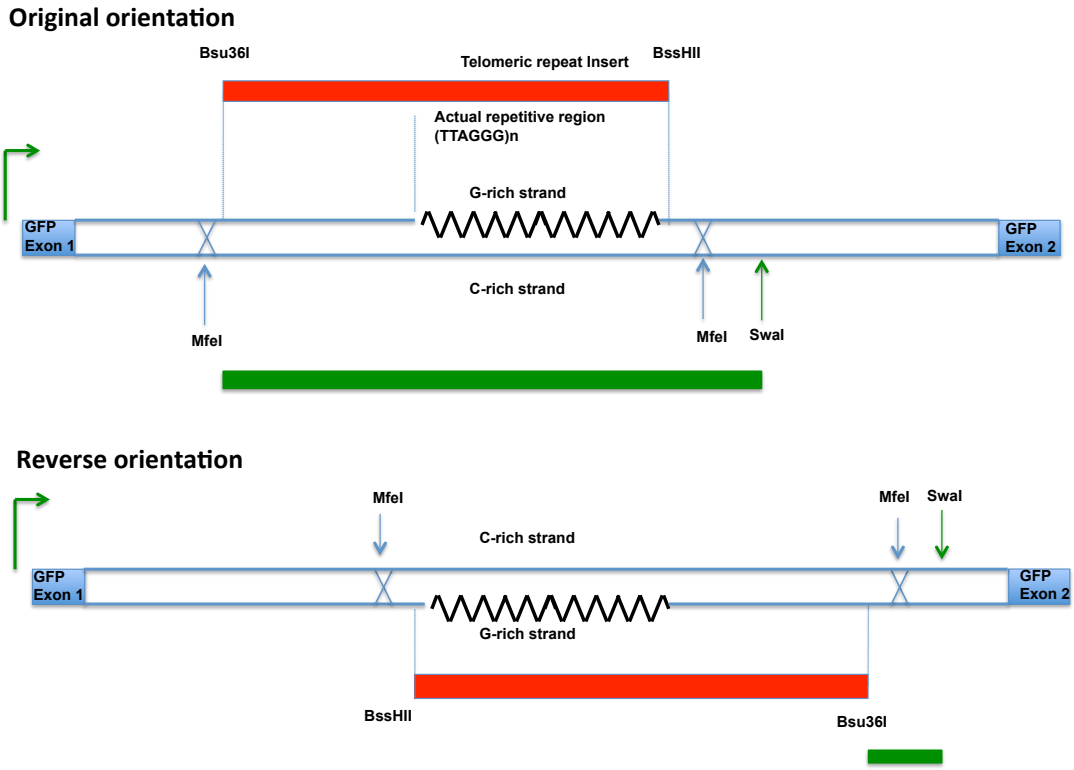

Figure 3.26 ***Schematic illustration showing how to check the inversion of the telomere repeat by double digestion with Bsu36I and SwaI****. Double digestion of the original (physiological) orientation results in a large fragment (green) (approximately 400bp or 500bp for Tel42 or Tel71, respectively) whereas the same digestion of the reverse orientation gives rise to a small fragment (12bp) (green).*

As shown in Figure 3.27A and B, clone Rtel71(2), (5), (6), (7), (8), (9), and (10) and clone Rtel42(1), (2), (5), (6), (8), (9) and (11) contain reverse orientation telomeric repeats as they produced no lower band compared to the original clone Tel71(5) and Tel42(5). This result was finally confirmed for clone Rtel71(2), (5), (6), (7), Rtel42 (1, (2), (9) and (11) by subsequent sequencing. Clone Rtel71(5) and Rtel42(1) (Figure

3.28 and Figure 3.29) were chosen for generating stable expression mammalian cell lines.

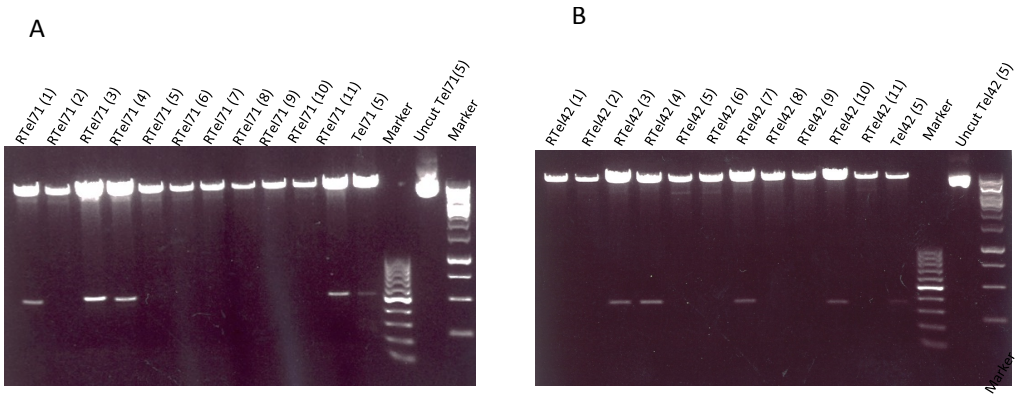

Figure 3.27 ***Double digest with Bsu36I and SwaI of plasmids isolated from the indicated bacterial clones to select for the ones with an inverted telomere repeat****. (A) Double digest of plasmids from Rtel71 clones. (B) Double digest of plasmids from Rtel42 clones. The marker is 100bp DNA ladder (NEB).*

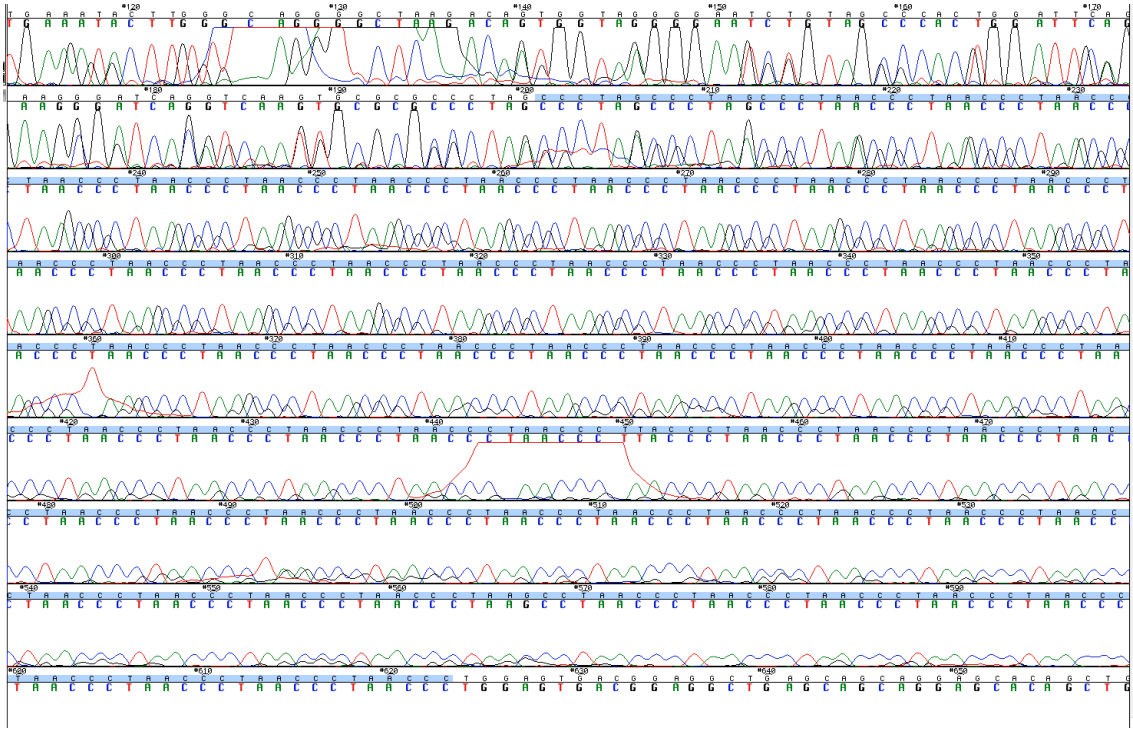

Figure 3.28 ***Sequencing result of a plasmid isolated from bacterial clone Rtel71(5) showing a reverse oriented telomere repeat of 71 units (CCCTAA)71****.*

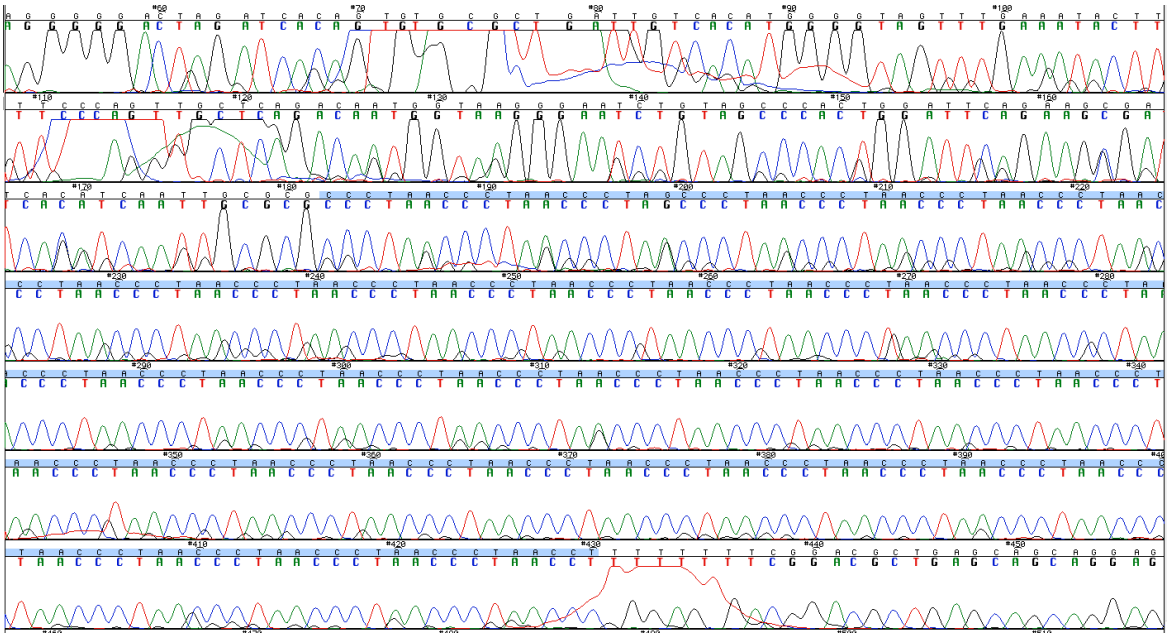

Figure 3.29 ***Sequencing result of a plasmid isolated from bacterial clone Rtel42(1) showing a reverse oriented telomere repeat of 42 units (CCCTAA)42.***

Figure 3.30 and Figure 3.31 show verification of the 293T-REx clones which were transfected with the chosen constructs and were subsequently selected by hygromycin resistance and zeocin sensitivity assays. All of the clones contain a specifically integrated cassette with a repeat of a corresponding size. The sequence and orientation of the repeat was confirmed by sequencing on clone Rtel71(3), Rtel71(4), Rtel71(5), Rtel42(6), Rtel42(7) and Rtel42(8).

3.2.3 Expression of the ectopic reporter gene

All of the 293T-REx clones generated containing ectopic repeats were confirmed by FACS analysis or Real-time PCR to have transcription switched on by addition of doxycyline (an analog of tetracycline with a similar mechanism of action). An example is shown on Figure 3.32.

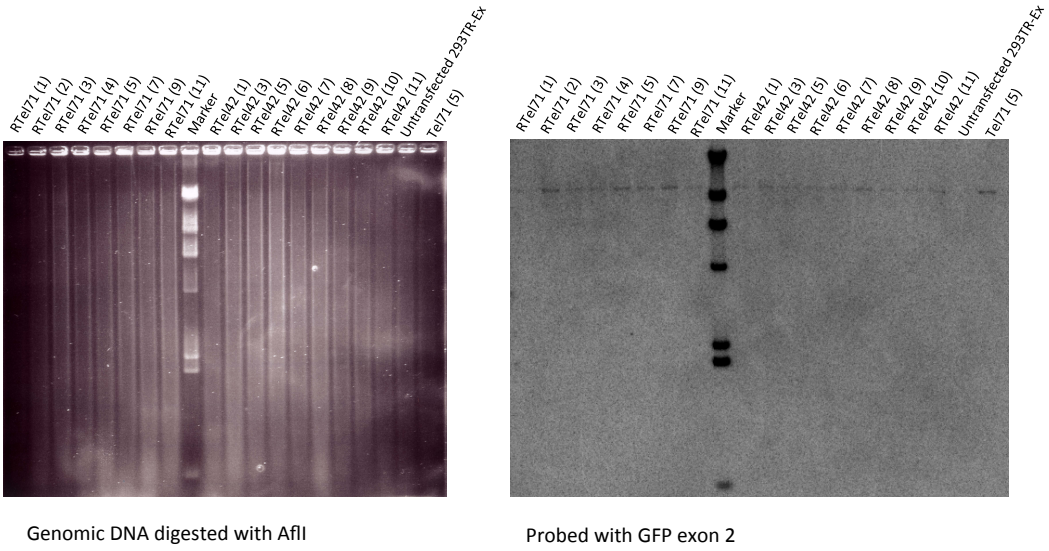

Figure 3.30 ***Southern Blot verifies single integration of reverse telomere repeat cassettes in corresponding clones****. Genomic DNA of the indicated clones was digested with AflI and then probed with P32 labelled GFP exon 2.*

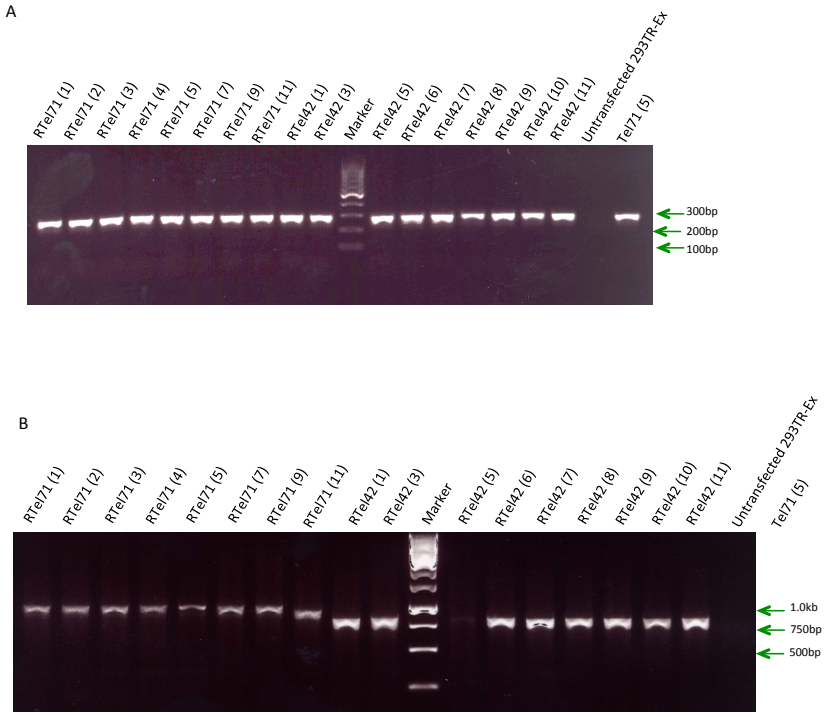

Figure 3.31 ***Verification of reverse telomere repeat containing clones****. (A) Confirmation of successful integration of ectopic reverse telomere constructs at the specific FRT site by PCR using the primer pair pSV40F/Hygromycin R. (B)*

Confirmation of the presence of ectopic reverse telomere repeat in the genome of indicated clones by PCR using the primer pair UpstreamIntF/DownstreamIntR. Untransfected 293T-REx and Tel71(5) serve as a negative control and a positive control, respectively.

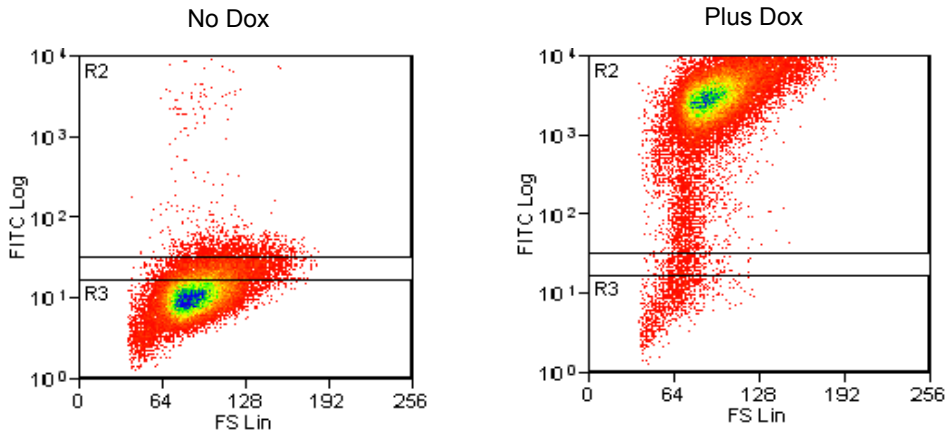

Figure 3.32 ***Expression of the ectopic reporter gene GFP*** *is switched on upon addition of doxycycline (1μg/mL) to the culture. FACS analysis shows an example of the switching in clone LB8R-5.*

3.3 DISCUSSION

In this chapter I have described the generation of stable expression mammalian cell lines carrying ectopic GC-rich tandem repeats. Success of the work in this first part of the project is vital for the subsequent parts as it provides materials for investigating ATRX binding at the repeats.

To my best knowledge, ψζVNTR has never been cloned into an ectopic location in mammalian cells. Human telomeres of different sizes, however, have been cloned into

a hamster-human hybrid cell line (500bp) (Farr, Fantes et al. 1991), Chinese hamster ovary cells (800bp) (Kilburn, Shea et al. 2001) and a human fibrosarcoma cell line HT1080 (1kb) (Itzhaki, Barnett et al. 1992), by random integration or through homologous recombination at specific sites.

As both ψζVNTR and telomeres are highly repetitive and GC-rich, manipulation of these during generation of the stable cell lines presented a number of difficulties. First of all, PCR amplification and sequencing analysis of these repeats are hampered by high melting temperatures and the probable formation of secondary structures. For ψζVNTR, in which GC content is 96%, addition of 5M Betaine and DMSO buffer (described in Chapter 2 Materials and Methods) to the PCR reactions provided a solution and I managed to produce good yields of specific PCR products (Figure 3.4). However, it was still impossible in my hand to sequence through the entire long ψζVNTR repeat. PCR amplification of telomeric repeats was difficult in a different aspect. Faststart taq DNA polymerase (Roche) was effective enough without supplementation with Betaine and DMSO buffer. However, because the forward primer for this PCR contains only telomere repeat, which could anneal anywhere along the telomere DNA template, the product size ranges from very small (70bp primer dimers) to large (a few kb, the size of a telomere). A particular challenge was that small products (below 500bp) were preferentially amplified, indicated by a strong signal at the bottom of the smear on the gel in Figure 3.20. One of my aims for this part was to generate a construct containing a 570bp telomere repeat, which is comprised of the same number of repeat units as the longest ψζVNTR obtained (about 95 units), so that they would be comparable when investigating ATRX binding.

However, after many attempts, the longest telomere repeat that I was able to clone was 71 units in length.

The second hurdle in making the stable expression cell lines was at the cloning step. In general, repetitive DNA sequences cloned in bacterial cells are highly prone to rearrangement, elimination and abbreviation. At the start of the project, DH5α *Escherichia coli* bacterial strain (Life Technologies), which has a mutation in the gene responsible for general recombination *recA1*, was used to clone ψζVNTR into pcDNA5/FRT/GFP. This is apparently sufficient to conserve the repeat sequence although there was an abbreviation of the repeat derived from sample N0908 (clone N2 on Figure 3.6C).

Cloning telomere repeats proved more challenging. At the start, vectors pSXneo(T2AG3)70, pSXneo(T2AG3)135 and pSXneo(T2AG3)270 (kindly provided by Prof. Duncan Baird, Cardiff University) containing the human telomeric repeat (T2AG3) of different sizes (indicated by the number outside of the brackets) were used as a source of telomeric repeat for engineering into the pcDNA5/FRT/GFP vector. The strategy was to isolate the repeat fragment by PCR amplification, followed by restriction enzyme digestion and ligation into the target vector. However, technical difficulties in obtaining a robust PCR product meant that this approach was not viable. Another plasmid pTELN, which was believed to carry 800bp of human telomeric repeat was obtained from Prof. Nancy Maizels. However, after sequencing the 800bp repeat fragment, it was apparent that the fragment was rearranged, probably during many rounds of propagation in bacterial cells. Several attempts were made to directly subclone the telomere sequence derived from a patient with a truncated

chromosome "healed" with telomere sequence. This was finally achieved using a PGEMT vector. However, subsequent transferring of the repeat fragment from pGEMT to pcDNA5/FRT/GFP was attempted several times without any success. In each attempt, the strategy was slightly modified, including optimizing the annealing temperature, changing the restriction enzymes or the ligation enzyme and using *E. coli* strains that are known for cloning highly repetitive DNA, to negate the problem.

Following many unsuccessful attempts using conventional cloning approaches, CloneEZ (GenScript) was employed. Unlike conventional cloning methods which rely on restriction enzyme digestion and ligation reaction catalyzed by T4 ligase, CloneEZ employs PCR and homologous recombination (Figure 3.20C). It is speculated that the telomere repeats might be more unstable in bacteria than the ψζVNTR despite the fact that the GC content of the former (50%) is lower than that of the latter (96%). Although Stbl2 competent cells (Life Technologies), a bacterial strain specifically designed for cloning unstable DNA, were used, several rounds of cloning failed at the transformation step resulting in no colonies. The failure could also be due to the ligase since CloneEZ, a ligase-free method, was effective.

Moreover, screening for colonies that successfully harbored the telomere repeats were a significant challenge because of a high background of colonies having the original pcDNA5/FRT/GFP which also contains an insert with approximately the same size as the telomere repeat insert. Colony lift assay, which is based on the Southern Blot technique, provided an excellent solution for an efficient high through-put screening.

After transfection of the repeat constructs into 293T-REx cells, clones that have formed on hygromycin containing medium were picked and they were then subjected to zeocin sensitivity and Beta-Gal staining assays to ensure a correct incorporation into the FRT site, according to the manufacturer. These assays were applied for screening the physiological oriented ψζVNTR constructs. However, I found that the use of both screening techniques was redundant given that the zeocin resistance gene and the beta galactosidase gene are of a fusion gene and hence inactivation of one, as the consequence of losing its promoter and the start codon after the integration event, should be effective for the other as well. Therefore, subsequently only zeocin sensitivity assay was carried out for clone screening of the other constructs.

Although Flp-In T-Rex system is straightforward, one caveat is that the FRT site is unmapped by the manufacturer (Invitrogen). It is, however, predicted to be placed in a transcriptionally active region of the genome. Not knowing the exact location of the integrated cassettes, it is impossible to estimate the size of the resulting fragment when digesting the genomic DNA of the clones with a unique interstitial cutter because the external site of the cutter is unknown. Nevertheless, the most important requisite was that there was only one band in each clone, regardless of size, indicating that the integration event occurred only once at a single site in the genome. The bands from different clones are expected to have the same size if the digested cassette fragments are from the same location in the genome. However, bands arising from clone N2S4, N1-2, LB8-8 and LB8-9 are at a different size than the ones from the other clones (Figure 3.11C), raising the question of whether the integration has happened at a different site in the genome rather than at the FRT site. PCR analysis using primer pair pSV40 F/Hygromycin R, however, ruled out this possibility (Figure

3.12). It is, therefore, unclear how these bands are at a different size. Nevertheless, both results from Southern Blot and FRT PCR demonstrated that the integration of the ψζVNTR cassette in all the selected clones was a single event at specific FRT site. This problem did not occur with the reverse orientation ψζVNTR clones or the telomere repeat clones.

In summary, despite all the discussed obstacles, stable expression clones containing ectopic ψζVNTR and telomere repeat of various sizes and in both orientations have been successfully generated to conduct an *in vivo* investigation of ATRX targeting that will be described in the next chapter.

4 ATRX binding at the ectopic GC-rich tandem repeats

4.1 INTRODUCTION

As discussed in the introduction to chapter 3, detailed knowledge about how the chromatin remodeler ATRX is recruited to its genomic location is at present lacking but this information would provide a great insight into the molecular functions of ATRX and help decode its mechanism of gene regulation. The previous chapter described the preparation of material for the work in this chapter: the generation of stable inducible expression mammalian cell clones carrying ectopic GC-rich tandem repeats. The experiments in this chapter will employ these clones to address three questions.

First, how does transcription influence ATRX targeting at the repeats? Using ChIP-on-chip and ChIP-seq on the 500kb terminal region of chromosome 16p in human erythroblasts, fibroblasts and Hep3B cells, ATRX has been shown to bind G-rich tandem repeats located in or near the genes whose expression is perturbed in ATR-X patients (Law, Lower et al. 2010). Interestingly, some ATRX targets seemed to be tissue-specific, such as the VNTR in the ψζ gene whereas other targets appeared consistently in all three cell types, such as NME4 and telomere repeats. Given that the NME4 and telomere repeats are transcriptionally active in all these cell types but the ψζVNTR is active only in erythroid cells (Chris Fisher, personal communication), one implication could be that transcription plays a role in ATRX recruitment to its targets. In support of this, our genome-wide analysis shows that a subset of ATRX targets, which are methylated CpG islands, are linked to transcription (Hsiao Voon, unpublished data). To elucidate this point, chromatin immunoprecipitation with an

ATRX antibody will be performed on the generated stable inducible expression clones containing either ectopic ψζVNTR or telomere repeats in active and inactive states of transcription.

The ψζVNTR is highly polymorphic. Interestingly, the study by Law and coworkers showed a significant correlation between the size of the tandem repeat in ATR-X patients and the severity of alpha thalassemia (reflecting the degree of alpha globin reduction) measured by the number of cells with HbH inclusion (Law, Lower et al. 2010). The larger the tandem repeat, the greater the reduction of alpha globin expression. This leads to the second question: how does the repeat size affect ATRX targeting? To this end, a ChIP assay was performed with different sizes of the ectopic repeats.

Finally, as discussed in chapter 1, a gel shift assay has shown that ATRX binds a non-B DNA structure called G-quadruplex (G4) that has the potential to form at sequences with a high GC content *in vitro* and *in vivo* (Law, Lower et al. 2010) (Gellert, Lipsett et al. 1962, Williamson, Raghuraman et al. 1989, Schaffitzel, Berger et al. 2001, Biffi, Tannahill et al. 2013, Biffi, Tannahill et al. 2014). In addition, GC-rich DNA has extensively been shown to adopt R-loops during transcription. Studies have shown that these two structures preferentially form when the non-template strand is G-rich (Reaban, Lebowitz et al. 1994, Daniels and Lieber 1995, Yu, Chedin et al. 2003) (Duquette, Handa et al. 2004, Ginno, Lott et al. 2012, Powell, Coulson et al. 2013). In order to determine if these secondary structures are targeted by ATRX, an obvious question is: does reversing the orientation of the repeat change the pattern of ATRX

binding. ATRX ChIP performed on clones with the repeats in both transcriptional orientations will help to answer this question.

4.2 RESULTS

4.2.1 ATRX presence in 293T-REx cell line

To study ATRX binding at the ectopic repeats in the generated clones, it is essential to make sure that the ATRX protein is expressed normally in the cell line of interest. The expression of ATRX in the 293T-REx cell line, derived from the parental cell line HEK293, has never been studied previously. After several genetic manipulations through random integration, including the introduction of the FRT site and the incorporation of the tetracycline repressor expression locus, there was a slight risk that the ATRX gene might have been disrupted. A Western Blot analysis was performed to check the presence of ATRX in the 293T-REx cell line. 50µg of the cell lysate was run in duplicate in a 4-12% Bis-Tris polyacrylamide gel. Proteins were electroblotted and probed with a polyclonal antibody against ATRX H300, which recognises the full-length form of ATRX (Figure 4.1A). Figure 4.1B shows that ATRX protein is indeed present in the 293T-REx cell line, compared to the knock-down cell line, validating it for investigation of ATRX binding which will be described in the following parts of this chapter.

4.2.2 Validation of Quantitative Chromatin Immunoprecipitation with ATRX antibody in the 293T-Rex cell line

Chromatin immunoprecipitation (ChIP) assay was employed to investigate ATRX targeting at the ectopic repeats. This method allows quantifying protein binding at a region of interest. Using polyclonal antibody, H300, specific to ATRX (Figure 4.1B), it was possible to pull down all the regions in the genome that are bound by the

protein. ATRX binding level at an investigating region was measured by Quantitative Real-time PCR (Q-PCR), using SYBR green dye, on the DNA material resulting from the ChIP. In order to demonstrate an enrichment of the protein at the region, comparisons were made with negative and positive controls which are reference regions known to not bind or bind to ATRX. These references are usually taken from previously published studies or from next-generation sequencing analysis of the ChIP material (ChIP-seq). However, to the date of this study, no ATRX ChIP in the 293T-REx cell line had been reported in the literature. Therefore, these controls had to be determined empirically.

ChIP was performed in the 293T-REx cells with antibody H300 against ATRX, and as a negative control for the specificity of the antibody a reaction with non-immunised rabbit IgG was performed in parallel. After that, immunoprecipitated DNA was recovered for Q-PCR. To gain an understanding of the ATRX binding profile in the 293T-REx cell line, I chose a number of regions that were previously demonstrated to be highly enriched for ATRX in other cell lines; these included the ribosomal DNA loci, 16p telomere and endogenous ψζVNTR (enriched in erythroid cells) (Law, Lower et al. 2010). Two other non-repetitive non GC-rich regions, GAPDH (intron 2) and DIST (also called RHBDF1, intron 4), that have been described as non-targets of ATRX (Law, Lower et al. 2010), were also tested. The enrichment level was plotted relative to the percentage of the input DNA. The transcribed region of rDNA (rDNAp4) shows an enrichment of 14 fold above GAPDH (Figure 4.2), which is consistent with the previous study in Erythroblast and HEP3B cell line (Law, Lower et al. 2010). The enrichment at the end of the ribosomal DNA transcribed region (rDNAp5) in 293T-REx, however, is only 2.5 fold above GAPDH while at the non-

transcribed region (rDNAp8) it is 5.5 fold above GAPDH (Figure 4.2). Moreover, the distribution of ATRX across the locus in 293T-REx is different from that in the other two cell lines. While the ATRX binding level at rDNAp5 is higher than at rDNAp8 in Erythroblasts and HEP3B, it is lower than that at rDNAp8 in 293T-REx (Figure 4.2). The telomere region of the 16p chromosome shows an enrichment of 11.25 fold above GAPDH, and endogenous ψζVNTR 3.5 fold. The level of binding at DIST is as low as at GAPDH (Figure 4.2).

16ptel and DIST were chosen to be a positive and a negative control for subsequent ATRX ChIP experiments, and these helped determine whether an experiment had worked properly. The ChIP assay is a lengthy procedure, involving many steps, and therefore many variations could occur during the process, which can account for variation in its efficiency. In order to minimize the effects of these variations and make fair comparisons between samples across different ATRX ChIP experiments, performed at different times, 16ptel was also chosen for normalization.

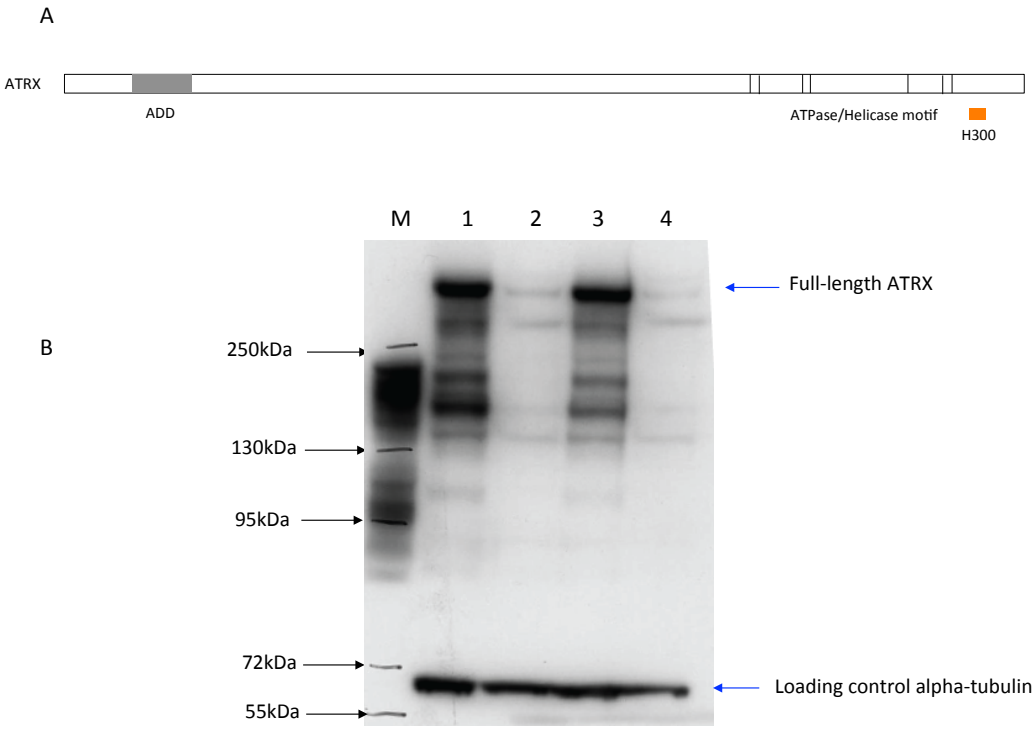

Figure 4.1 ***ATRX presence in the 293T-REx cell line****. (A) Location of epitopes for the polyclonal antibody H300 (orange block) that recognises the full-length ATRX (B) Western Blot analysis with H300 shows ATRX expression in wild-type 293T-REx cell line (Lane 1 and 3) compared to the ATRX knock-down 293T-REx cell line (Lane 2 and 4). The loading control is alpha-tubulin. Lane M: PageRuler Prestained protein ladder. Note that the H300 antibody is cross-hybridised with the marker proteins in the ladder.*

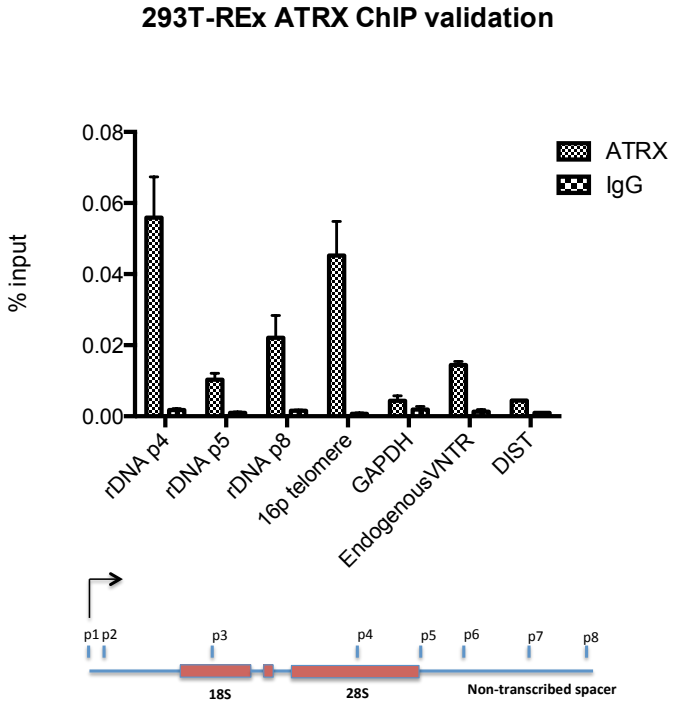

Figure 4.2 ***ATRX ChIP validation in the 293T-Rex cell line****. QPCR analysis of ATRX ChIP in the host cell line 293T-REx to validate the protocol and the antibody H300. The enrichment of ATRX was quantitated at rDNA using amplicon number 4, 5 and 8, at the telomere region of chromosome 16, GAPDH, endogenous ψζVNTR and DIST. The graph was plotted based on 3 independent ChIP experiments. Q-PCR for the amplicon rDNAp4, rDNAp5, 16ptel and GAPDH was carried out in all 3 ChIPs. Q-PCR for the amplicon rDNAp8 and endogenous VNTR was performed in 2 ChIPs; and for the amplicon DIST ChIP. The level of the enrichment is expressed as % input.*

4.2.3 ATRX binding at the ectopic ψζVNTR

4.2.3.1 ATRX binding at the ectopic ψζVNTR in the physiological orientation

In the physiological orientation, the ectopic ψζVNTRs are oriented such that the G-rich strand is the non-template strand, which means when the GFP gene is transcribed this G-rich strand is not the template for RNA polymerase II, and the nascent RNA transcript will have the same sequence as this G-rich strand. This is the orientation of the repeat in its natural genomic context in the ψζ gene.

To investigate ATRX targeting at the ectopic ψζVNTR, ChIP with antibody H300 against ATRX was carried out for clones carrying VNTRs of different sizes ranging from 0 to 490bp as indicated in Figure Figure 4.3A. These are clones generated by introducing ψζVNTR into the 293T-REx genome, which was described in detail in chapter 3. To account for clonal variation, ChIP was performed with at least two clones for each size of the VNTR. NoVNTR corresponds to clones N2S(4), N2S(5) and N2S(9); 140bp VNTR corresponds to clones N2(1) and N2(2); 240bp VNTR to clones N1(2), N1(4) and N1(5); 490bp VNTR to clones LB8(8), LB8(7) and LB8(9). However, due to a technical problem, for 390bp VNTR only one clone, N3(2), was used for ChIP although the experiment was repeated on two independent occasions.

For each clone, cells were grown in both the presence and absence of doxycycline to repress or induce GFP expression, respectively. The cells were induced for 48 hours before being harvested for ATRX ChIP. ChIP procedure and subsequent Q-PCR analysis of the ChIP DNA material were performed for the uninduced and induced states in parallel for each clone unless stated otherwise. Due to the repetitive nature of the VNTR, for the detection of ATRX, two Q-PCR amplicons, namely

DownstreamVNTR1 and DownstreamVNTR3 were designed in the flanking region rather than within the repeat (Figure 4.3C). The Downstream VNTR3 amplicon lies within the Downstream VNTR1 amplicon. Both are located downstream of the repeat insert, and are about 260bp downstream of the last G cluster in the actual repetitive region (Figure 3.3).

The ATRX enrichment seen at DownstreamVNTR1 and DownstreamVNTR3 for each length of the ψζVNTR are broadly similar (Figure 4.3A and B). When the gene was inactive, ATRX was enriched at approximately 3 to 5 fold above the background control DIST at all of the different sizes of the ψζVNTR. However, these enrichment levels are relatively low, about 0.3 to 0.5 fold, compared to the positive control 16ptel (Figure 4.3A). There is no significant difference in binding levels between different VNTR sizes, including the NoVNTR.

Interestingly, when gene expression was induced, the level of ATRX binding rose at all the different sizes of the VNTR, except for the NoVNTR, suggesting that transcription has a positive effect on ATRX binding to the repeat, and that this effect is attributable to the presence of the repeat. More strikingly, the degree of the increase appears to be directionally proportional to the length of the repeat, by approximately 1.3 folds at the 140bp VNTR, 2.5 fold at the 240bp VNTR, 3 fold at the 390bp VNTR and most dramatically 6-8 fold at the 490bp VNTR compared to the corresponding un-induced condition (Figure 4.3A and B). Altogether, this indicates that ATRX is recruited to the repeats in a transcription and size dependent manner.

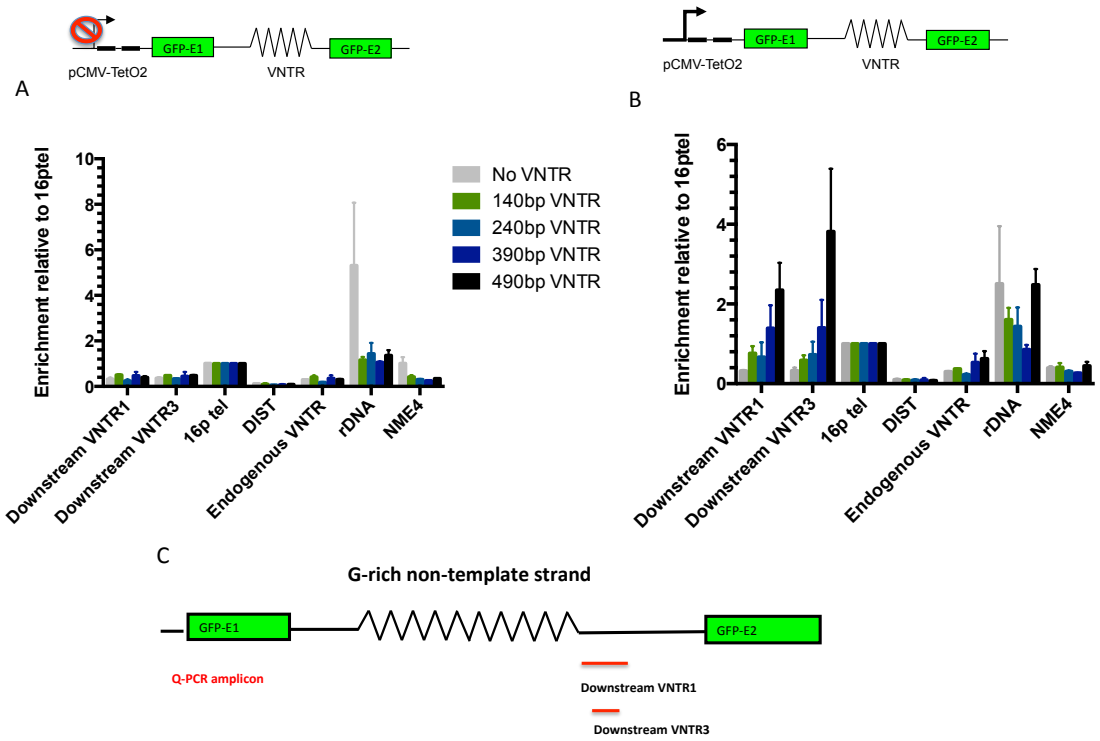

Figure 4.3 ***ATRX binding in the clones containing ectopic ψζVNTRs in the physiological orientation****. ATRX ChIP Q-PCR in clones carrying ψζVNTR of different sizes when (A) gene expression is repressed, and (B) when the gene is switched on by addition of doxycycline (1μg/mL). Enrichment for each amplicon was quantified by Q-PCR, normalising to %input and is expressed relative to 16ptel. Data bars were plotted as the mean of two independent experiments for the 140bp VNTR and 390bp VNTR; of three independent experiments for the NoVNTR and the 240bp VNTR; and of four independent experiments for the 490bp VNTR. Error bars represent standard error of the mean (S.E.M) (C) Diagram shows locations of the two amplicons that were used for Q-PCR to quantify the level of ATRX binding at the ectopic ψζVNTR. The Downstream VNTR3 amplicon lies within the Downstream VNTR1.*

To control for the effect of the drug doxycycline, the NME4 repeat, endogenous ψζVNTR and ribosomal DNA (Figure 4.3A and B), which are all GC-rich repeats, were included in all the analyses. As expected, these control regions were not

consistently changed upon induction between the clones, and therefore I concluded that any difference simply reflects technical variations.

It is worth noting that when the ectopic gene is inactive, all the clones of different VNTR sizes, including the NoVNTR, displayed a broadly similar level of ATRX binding at the ectopic VNTR relative to that at the endogenous VNTR (about 3-5 fold above the background DIST) (Figure 4.4). This suggests that the flanking sequence (Figure 4.5) that is shared among the NoVNTR, the ectopic VNTR and the endogenous VNTR may recruit low levels of ATRX.

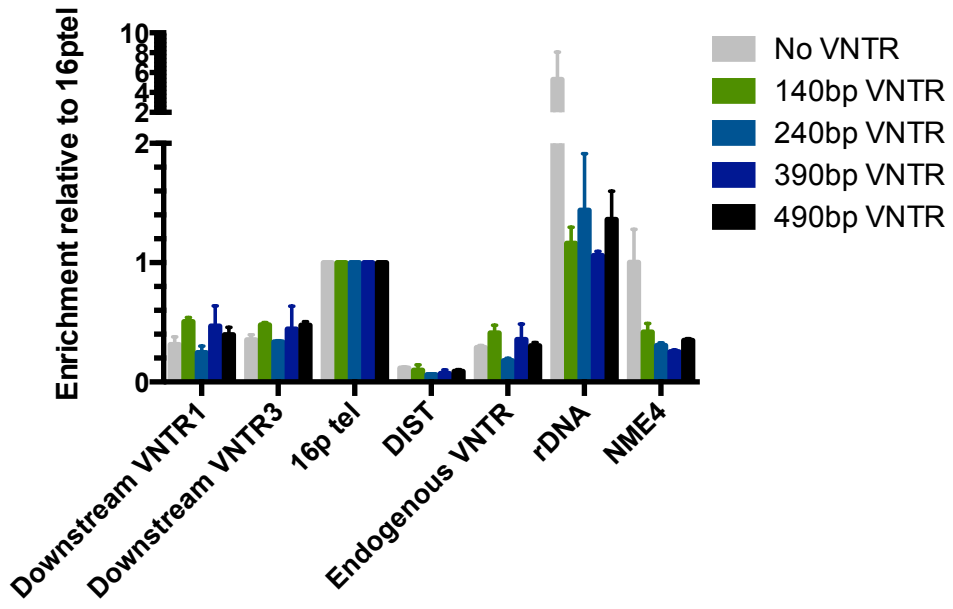

Figure 4.4 ***A zoom-in view of figure 4.3A****. The Y axis is split into segments to give a better view of the bottom part of the graph where most of the data bars are aligned within.*

CCTGAGGCAAGACCTACTTCCCGCACTTCGACCTGCACCCGGGGTCCGCGCAGTTGCGCGCT
TCCCCGCCGACTTCACGGCCGAGGCCCACGCCGCCTGGGCCAAGTTCCTATCGGTCGTATCC
TCTGTCCTGACCGAGAAGTACCGCTGAGCGCCGCCTCCGGGACCCCCAGGACAGGCTGCGG
CCCCTCCCCTGCCCTTCACCCTCCCACAGTTCCTGCCCTGACTCCAATAAATGGATGAGGACG
GAGCGATCTGGGCTCTGTGTTCTCAGTATTGGAGGGAAGGAACGCGT

Figure 4.5 ***The non-repetitive flanking sequence contains many C-runs****. Sequence of the No VNTR construct in which the G-rich segment has been removed from the original insert, between the Bsu36I (blue) and MluI (blue) sites, by using BssHII (green), leaving a flanking region with C-runs (red).*

4.2.3.2 ATRX binding at the ectopic ψζVNTR of the reverse orientation

Next, I addressed the question as to how ATRX binding at the repeat is affected by the transcription orientation of the G-rich strand. To this end, ATRX ChIP and Q-PCR analysis were performed as described above for the clones carrying ectopic ψζVNTR in the reverse orientation in which the G-rich strand is the template strand during transcription. The procedure was carried out for each of the clones N1R(4), N1R(5) containing 240bp reverse VNTR (rVNTR), clone LB8R(1), LB8R(2), LB8R(3) and LB8R(5) containing 490bp rVNTR in the induced and un-induced state.

Q-PCR analysis of immunoprecipitated DNA by antibody H300 shows that regardless of the reverse orientation, ATRX enrichment at the repeat still increases upon transcription induction (Figure 4.6A and B), suggesting that the orientation of the G-rich strand may not influence ATRX recruitment to the repeat. However, it appears that the increase in the binding at the long repeat (490bp VNTR) upon activation of expression is less than that in the physiological orientation: approximately 3 fold compared with 6-8 fold, respectively. Moreover, as in the physiological orientation, it seems that the enrichment level of ATRX at the reverse oriented VNTR also

increases, albeit less dramatically, with its size (Figure 4.6B): going from 240bp to 490bp the enrichment increases by 3.5-5.5 fold in the physiological orientation whereas in the reverse orientation it rises only by 1.5 fold.

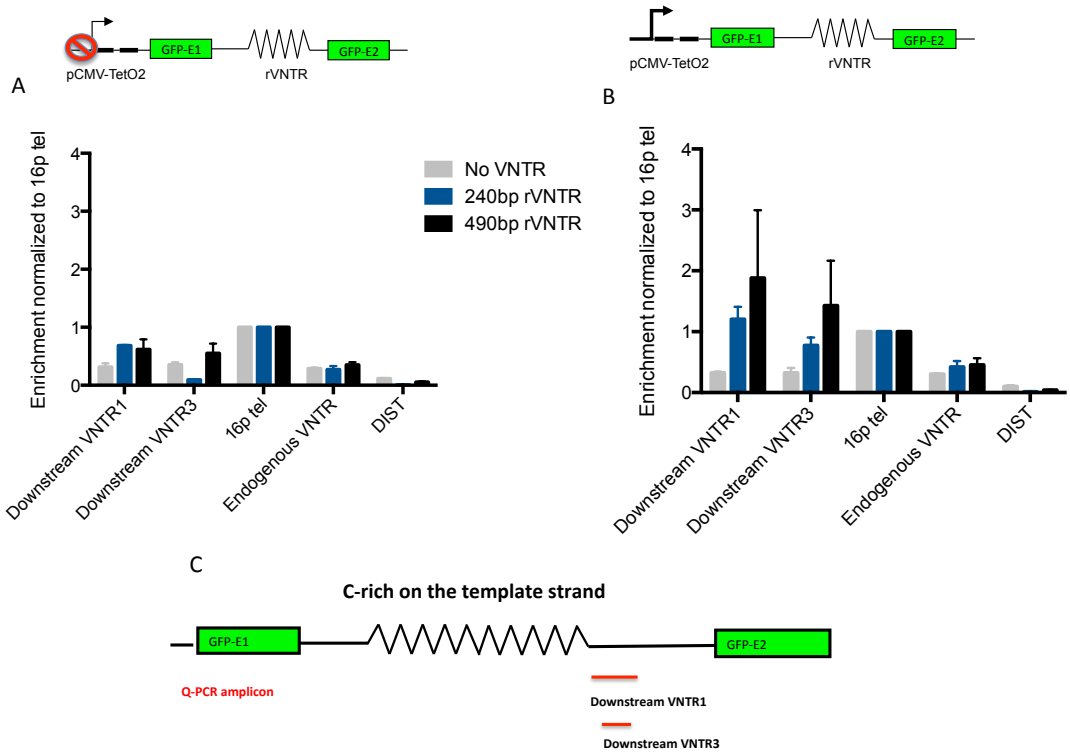

Figure 4.6 ***ATRX binding in the clones containing ectopic ψζVNTRs of reverse orientation****. ATRX ChIP Q-PCR in clones carrying reverse ψζVNTRs of different sizes when (A) gene expression was repressed, and (B) when the gene was switched on upon addition of doxycycline (1μg/mL). Enrichment at each amplicon was quantified by Q-PCR, normalising to %input and is expressed relative to 16ptel. Data bars were plotted as the mean of two independent experiments for 240bp VNTR; of three independent experiments for NoVNTR; and of four independent experiments for 490bp VNTR. Error bars represent standard error of the mean (SEM) (C) Diagram shows the locations of two amplicons that were used for Q-PCR quantifying the level of ATRX binding at the ectopic ψζVNTR. The Downstream VNTR3 amplicon lies within the Downstream VNTR1.*

4.2.4 ATRX binding at the ectopic telomere repeat

As discussed in the previous chapters, human telomere repeats, with TTAGGG as the repeating unit sequence, are a well-characterised target of ATRX. They are different from the ψζVNTR repeat (GCGGG)n in primary sequence, but share common features that make them prone to form similar secondary structures, including a high GC content, positive GC skew and G-clustering spaced by non-guanine bases in tandem repeats. I sought to address the question as to whether ATRX would interact with a telomere repeat in an ectopic genomic location and if so would the interaction be dependent on transcription, repeat size and the orientation of the G-rich strand. Similar to the ψζVNTR repeat, ChIP with the ATRX antibody H300, followed by SYBR green quantitative PCR, was used to look for ATRX enrichment.

4.2.4.1 ATRX binding at the ectopic telomere repeat in the physiological orientation

Telomeres are transcribed by DNA-dependent RNA polymerase II into telomere repeat-containing RNA (TERRA) (Solovei, Gaginskaya et al. 1994) (Azzalin, Reichenbach et al. 2007) (Schoeftner and Blasco 2008). Telomere transcription is mainly initiated at CpG islands located at approximately half of the subtelomeres (Nergadze, Farnung et al. 2009), using the C-rich strand as the template and is always directed towards the end of a chromosome (Solovei, Gaginskaya et al. 1994). This is called the physiological orientation of telomere repeats. This section describes my investigation on the interaction of ATRX with the ectopic telomere repeats that are constructed in the 293T-REx genome in the native orientation with regard to transcription.

To examine how the repeat size plays a role in determining ATRX targeting at the ectopic telomere repeats, the 293T-REx clones containing 42
and 71 repeating units of telomere {Tel71 or (TTAGGG)71} were used.

To determine whether the transcription process is a regulatory factor in ATRX targeting at the ectopic telomere repeats, ChIP was performed with the ATRX antibody H300 as described above, for the telomere repeat clones in un-induced and induced states. To induce gene expression, doxcycline was added to the culture to a final concentration of 1μg/mL, and after 24 hours of induction, the cells were harvested for ChIP. To account for clonal and technical variation, at least three independent ChIP experiments were performed on at least two clones for each size of ectopic telomere repeat: clones Tel42(5) and Tel42(6) were chosen for the length of 42 units, clone Tel71(5), Tel71(6) and Tel71(7) were chosen for 71 units. To enable quantitation of ATRX binding to the ectopic telomere repeats by ChIP, two Q-PCR amplicons, namely U2 and D1, were first designed, located at the immediate 5' end and 3' end of the repeat (Figure 4.7C). ATRX enrichment was calculated, as before, relative to % input and normalized to 16ptel.

Similar to the ψζVNTR, the ectopic telomere repeats also recruited more ATRX upon the activation of gene expression (Figure 4.7A and B) and the longer the repeat, the more dramatic the increase in ATRX recruitment: 3 fold at the D1 amplicon in the (TTAGGG)71 clone compared with 1.5 fold in the (TTAGGG)42 clone. Such an increase did not occur at other endogenous GC-rich repeats (rDNAp4, NME4 and Endogenous ψζVNTR) in the genome, indicating that doxycyline only acts on the

ectopic locus. Surprisingly, no difference in enrichment was detected upstream of the repeats using the U2 amplicon in both clones.

To extend the investigation of ATRX binding further up and down the repeat, more Q-PCR amplicons were designed, including U1 situated 100bp from U2, D2 300bp from D1, D3 at the start of GFP exon 2 which is 550bp from D1, D4 at the polyA signal which is 950bp from D1 and D5 immediately before the FRT site which is 1350bp from D1 (Figure 4.7C). Unexpectedly, ATRX binding signal did not drop but was maintained at a high level downstream of the repeat in the induced condition, but not in the un-induced condition. It seems that the further downstream the more the increase in ATRX binding, by 3 fold at D1, 4 fold at D2, 4.6 fold at D3 and 5.5 fold at the D4 amplicon over the un-induced ones (Figure 4.7B). In accordance with the previous result, this effect is strongly proportionate to the increased repeat size (Figure 4.7A and B). Interestingly, the degree of the increase dropped substantially at the D5 amplicon, where transcription might be terminated, indicating that ATRX recruitment upon transcriptional activation might be co-transcriptional. More intriguingly, upstream of the repeat, represented by the U1 and U2 amplicons, there was no increase in ATRX enrichment upon transcription induction in both clones tested (Figure 4.7A and B), leading to an asymmetrical ATRX distribution across the ectopic telomere repeats.

4.2.4.2 ATRX binding at the ectopic telomere repeat in the reverse orientation

To understand how transcriptional orientation of the G-rich strand affects ATRX recruitment to GC-rich repeats, ChIP with ATRX antibody H300 was performed in the clones carrying a 71 unit telomere repeat in the reverse orientation (CCCAAT)71.

In this orientation, the G-rich strand is used as the template for transcription when doxycycline is added to the cell culture. ATRX enrichment at these reverse repeats was quantitated by Q-PCR analyses using the same amplicons that were used for the physiological orientation (Figure 4.8C). Very interestingly, the increase in ATRX enrichment upon transcription that was observed at the physiological oriented repeats no longer occurs when the orientation was reversed (Figure 4.8A and B). This suggests that ATRX recruitment to the ectopic telomere repeats is dependent on the transcription orientation of the G-rich strand.

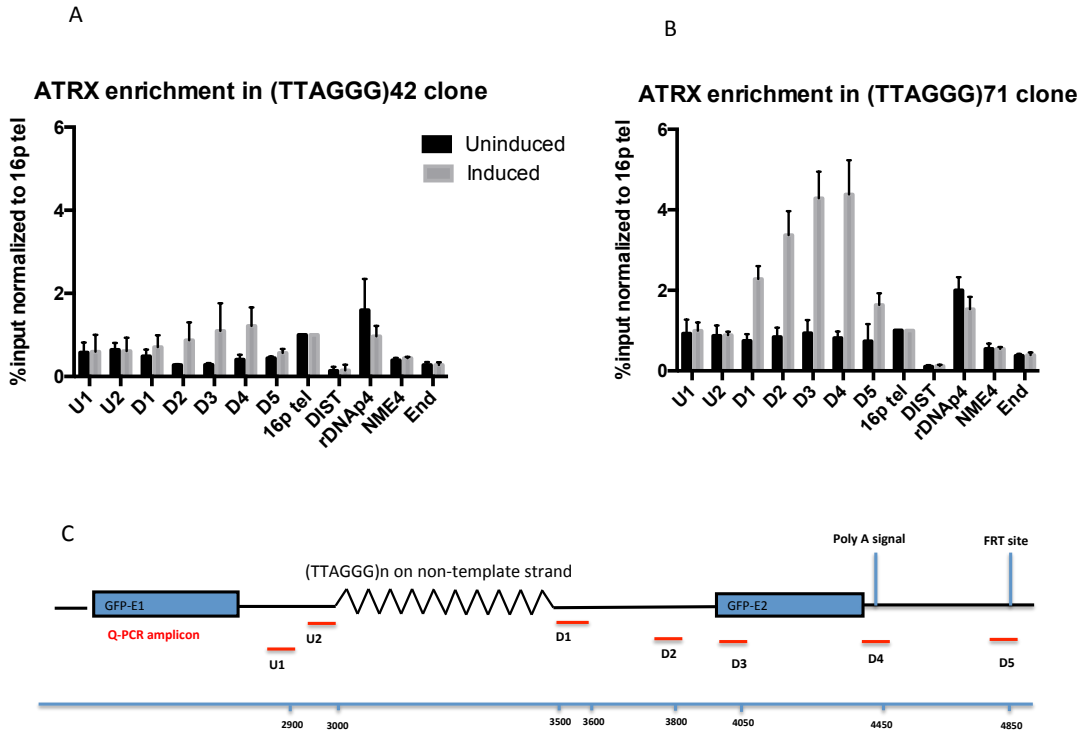

Figure 4.7 ***ATRX binding in the clones containing ectopic telomere repeats of physiological orientation (TTAGGG)n****. ATRX ChIP Q-PCR in (A) clones carrying a telomere repeat of 42 units, and (B) clones carrying a telomere repeat of 71 units in two conditions: when gene expression was repressed or switched on by addition of Doxycycline (1µg/mL). Enrichment for each amplicon was quantified by Q-PCR, normalising to %input and is expressed relative to 16ptel. Data bars on graph A were*

plotted as the mean of three independent experiments of three Tel42 clones. Data bars on graph B were plotted as the mean of four independent experiments of three Tel71 clones. Error bars represent standard error of the mean (SEM). (C) Diagram shows locations of the Q-PCR amplicons and their coordinates on the construct pcDNA5/FRT/GFP. The diagram is not drawn to scale.

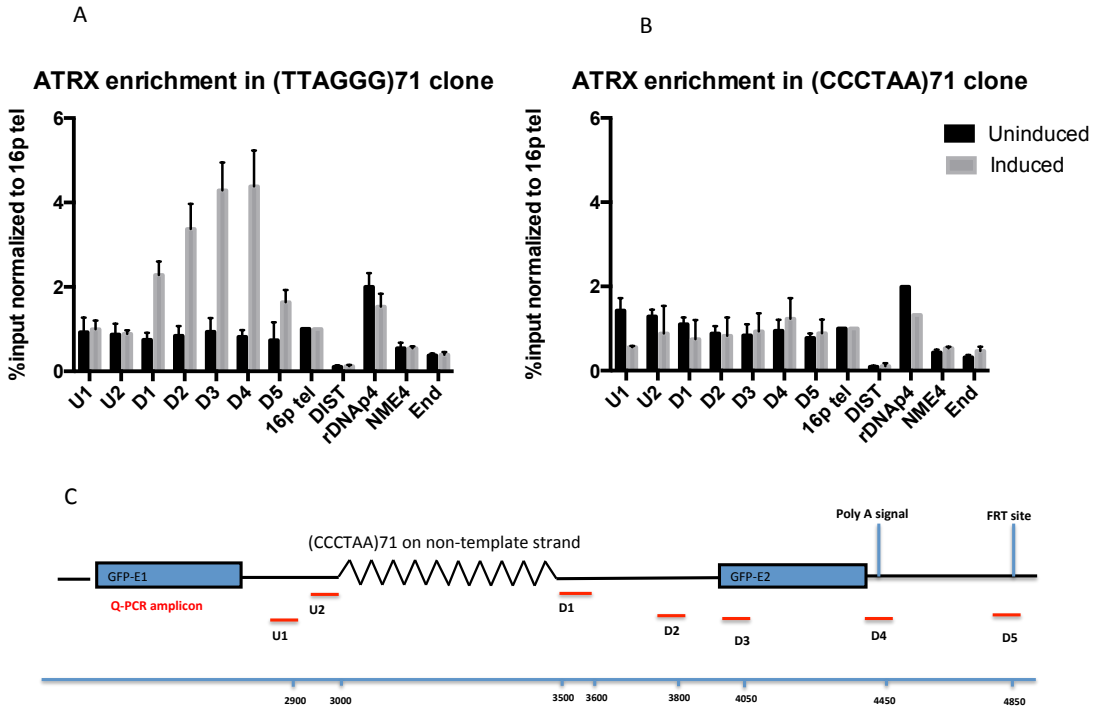

Figure 4.8 ***ATRX enrichment upon at the repeats is dependent on G skew on the non-template strand****. (A) the same figure as figure 4.7B. (B) Q-PCR with ATRX ChIP DNA from clones carrying 71 unit telomere repeats of the reverse orientation (the non-template strand is C-rich) in two conditions: when gene expression was repressed or switched on by the addition of doxycycline (1µg/mL). Enrichment for each amplicon was quantified by Q-PCR, normalising to % input and is expressed relative to 16ptel. Data bars were plotted as the mean of three independent experiments from two clones. Error bars represent standard error of the mean (SEM). (C) Diagram shows locations of the Q-PCR amplicons and their coordinates on the construct pcDNA5/FRT/GFP. The diagram is not drawn to scale.*

4.3 DISCUSSION

4.3.1 The ectopic GC-rich tandem repeats recruit ATRX independently from genomic context

Previous studies reported that ATRX preferentially bound GC-rich tandem repeats in the genome (Law, Lower et al. 2010). However, it was unclear as to whether this association of ATRX was guided by the intrinsic features of the repeats or by their surrounding genomic context. By showing in both cases, after taking out two GC-rich tandem repeats, ψζVNTR and telomere repeat, from their endogenous environments and putting them in a random genomic location, ATRX is still recruited to these repeats, the work described in this chapter demonstrates that GC-rich tandem repeats are sufficient to recruit ATRX independently from genomic context.

4.3.2 ATRX targeting at the ectopic GC-rich tandem repeats is facilitated by their transcriptional activity

At both ectopic ψζVNTR and telomere repeats, ATRX enrichment increases upon transcription (Figure 4.3 and Figure 4.7), suggesting that transcriptional activity facilitates ATRX binding to the repeats.

This novel finding might explain why some ATRX targets are tissue-specific and the link between ATRX enrichment with expression of its targets in different cell lines (Law, Lower et al. 2010).

Transcription is a complex process which involves many changes in transcription factors, histone modifications, histone variants and nucleosome occupancy, any of

which might serve as targeting signals for chromatin remodelers (see review by (Erdel, Krug et al. 2011)). For example, it has been shown in yeast that SWI/SNF can be recruited to gene promoters by DNA-binding transcriptional activators, independently from RNA polII holoenzyme and TBP (Yudkovsky, Logie et al. 1999). Induction of transcription of the Hsp70 gene leads to the recruitment of Xnp, the *Drosophila* homolog of ATRX, to nucleosome-depleted regions (Schneiderman, Orsi et al. 2012). Previous studies have demonstrated that the ADD domain of ATRX locates to pericentric heterochromatin via "reading" trimethylated lysine 9 on histone H3 (H3K9me3) (Dhayalan, Tamas et al. 2011, Eustermann, Yang et al. 2011, Iwase, Xiang et al. 2011). It is possible that an increase in H3K9me3 at the ectopic repeats upon transcription induction leads to the increase in ATRX enrichment. This possibility will be explored in the next chapter.

A common view is that chromatin remodeling removes physical barriers for gene transcription by changing chromatin conformation, repositioning nucleosomes and allowing the underlying DNA to be more accessible to transcription factors. Based on this view, it is rather contradictory that the targeting of the chromatin remodeler ATRX is signaled by transcription because in this model it is supposed to precede transcription. Furthermore, modification of histone tails is considered as part of the remodeling process (Luo and Dean 1999), and incorporating histone variants in a replication-independent manner may require chromatin remodelers (Li, Pattenden et al. 2005, Wu, Alami et al. 2005). It is possible that in our experimental system, ATRX is not required to initiate transcription but may play an essential role at the repeat to maintain transcriptional activity. Supporting this hypothesis, transcription is activated in the NoVNTR control upon addition of doxycycline despite no increase in ATRX

binding. This implies that it is the interplay between transcription and repetitive DNA that leads to recruitment of ATRX.

As discussed in Chapter 1, the literature suggests that GC-rich repeats can adopt non-canonical DNA structures such as G4 during cellular processes in which chromatin is temporarily loosened and the DNA duplex is transiently separated into two single strands (Duquette, Handa et al. 2004). Transcription facilitates the formation of G4 through inducing negative superhelicity (Sun and Hurley 2009). In addition, R-loops can be formed co-transcriptionally at sequences with high GC content and high transcription frequency (Daniels and Lieber 1995, Yu, Chedin et al. 2003, Luke, Panza et al. 2008, Aguilera and Garcia-Muse 2012, Chan, Aristizabal et al. 2014, Haeusler, Donnelly et al. 2014). It has been reported that R-loops play a role in chromatin remodeling at CpG island promoters, by preventing DNA methylation by the DNA methyltransferase DNMT3B1 (Ginno, Lott et al. 2012). It is tempting to speculate that ATRX utilizes such secondary structures as a targeting signal. Interaction between ATRX and pre-folded G-quadruplexes *in vitro* has been documented (Law, Lower et al. 2010). However, whether these abnormal structures form at the investigating ectopic repeats *in vivo* and whether ATRX actually interacts with them in the cells remain to be elucidated and will be discussed in the following chapter.

It is worth noting that there is a significant enrichment of ATRX at both the ψζVNTR and telomere repeat compared to the background DIST in the un-induced condition (Figure 4.3 and Figure 4.7). Furthermore, the same level of enrichment was also obtained at the NoVNTR control, suggesting that the flanking sequence (Figure 4.5)

in this construct recruits low levels of ATRX. It is possible that there was a low level of transcription that was leaking from the repressed pCMV TetO2 promoter, which could be sufficient to induce changes in the targeting signal leading to ATRX recruitment. In addition, the flanking sequence in the NoVNTR construct contains a number of C clusters, which could potentially form non-B form DNA structures which bind ATRX during leaky transcription. Supporting this argument, the ATRX enrichment at the flanking region (the average of the DownstreamVNTR1 and DownstreamVNTR3 grey bars in Figure 4.4) is approximately 0.4 fold over that at the 16p telomere positive control in the NoVNTR clones, whereas the enrichment level at the reverse oriented telomeric repeat, which contains more C-runs, is comparable to that at the positive control (the average of the D1-D5 dark bars in Figure 4.8B) in the (CCCTAA)71 clones. Alternatively, other DNA processes such as replication also promote G-quadruplex formation, which might explain the background enrichment of ATRX in the un-induced condition. This can be tested using transcription inhibitors.

4.3.3 The efficiency of ATRX recruitment to GC-rich tandem repeats is proportionate to the repeat size

Law and colleagues established a correlation between the size of the ψζVNTR, which is associated with the α-globin locus, and the degree of the gene repression in the absence of ATRX. The experiments in this chapter have demonstrated the correlation between the size of GC-rich repeats and the level of ATRX recruitment. In both cases, at the ψζVNTR and telomere repeats, the longer the repeat the more efficiently it recruits ATRX and moreover, this size effect occurs only when transcription is active but not when it is inactive (Figure 4.3 and Figure 4.7).

This novel finding suggests that an intrinsic feature of the repeats, such as the number of repeating units may exert a dosage effect on the affinity to ATRX. It has been reported that the G-rich minisatellite CEB1 is unstable in the absence of PIF1, an enzyme which resolves G4, with instability increasing with the size of the repeat array (Ribeyre, Lopes et al. 2009); the probability of G4 formation in a G-rich repeat would be expected to increase with its size. In other studies in which G-rich repeats form R-loop structures during transcription, a certain length of G-rich sequence is required to cause gene silencing because sufficient length of the repeat sequence in the nascent RNA transcripts is needed to reach back and hybridise with the DNA template (Colak, Zaninovic et al. 2014, Haeusler, Donnelly et al. 2014). It is hence tempting to postulate that the longer the repeat the more likely it is to form a non-canonical DNA structure, which consequently attracts more ATRX.

4.3.4 Transcription orientation of the G-rich strand influences ATRX targeting to GC-rich repeats

The effect of the transcription orientation in the ψζVNTR repeats was unclear possibly due to a mix of G and C clusters on the same strand (Figure 4.9). Reversing this would lead to runs of Gs which might form secondary structures. The ATRX binding level in the physiological orientation in which the number of G-runs is predominant over the number of C-runs appears to be higher than in the reverse orientation (Figure 4.3B and Figure 4.6B). At the ectopic telomere repeats, where G-skew is maximal (equal to 1) and there is no mix of C clusters (Figure 4.10), the increase in ATRX recruitment to the inverted repeats upon transcription was completely abolished (Figure 4.8).

CCTGAGGCAAGACCTACTTCCCGCACTTCGACCTGCACCCGGGGTCCGCGCAGTTGCGCGCG
CACGGCTCCAAGGTGGTGGCCGCCGTGGGCGACGCGGTGAAGAGCATCGACGACATCGGC
GGCGCCCTGTCCAAGCTGAGCGAGCTGCACGCCTACATCCTGCGCGTGGACCCGGTCAACTT
CAAGGTGCGCGGGGCGCGGTGCGGGCGGGGCGGGGCGGGGCCGCGGGGCGGGCGGGGC
CGCGGGGCGGGGTCGCGGGGCGGGGCGGGGTGGGGTCGCGGGGCGGGGCGGGGTCGCG
GGGCGGGGCGGGGCGGGGCGGGGCGGGCGGGGCGGCCGGGGCCCGGCGGGGCGGGGC
GGGGCGGGGAGGGGCTGGGCGGGGCGGGGCGCGGGGCGGGGCGGGCCGGGCCGGGGC
GGGGTCGCGGGGCGGGGTCGCGGGGCGGGGCGCGGGGCGGGGCGGGGCGGGGTGGGG
TCGCGGGGCGGGGCCCGGGCTAGGCCCCGCCCCCGCACTGAGCCGCCCCCGCCCCCAGCTC
CTGTCCCACTGCCTGCTGGTCACCCTGGCCGCGCGCTTCCCCGCCGACTTCACGGCCGAGGC
CCACGCCGCCTGGGCCAAGTTCCTATCGGTCGTATCCTCTGTCCTGACCGAGAAGTACCGCT
GAGCGCCGCCTCCGGGACCCCCAGGACAGGCTGCGGCCCCTCCCCTGCCCTTCACCCTCCCA
CAGTTCCTGCCCTGACTCCAATAAATGGATGAGGACGGAGCGATCTGGGCTCTGTGTTCTCA
GTATTGGAGGGAAGGAACGCGT

Figure 4.9 ***Mix of G clusters (blue) and C clusters (green) on the same strand of the Ψζ VNTR insert****, which is cloned into the construct pcDNA5/FRT/GFP. The sequence between two restriction sites BssHII (orange) is inverted in the reverse orientation clones.*

CCTGAGGTGCACGCCCAGATCACGGATGTGTCGGGAACAGCTGTGCTCCTGCTGCTCAG
CGTCCGTCACTCCAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGCGCGC

Figure 4.10 ***Sequence of the telomere repeat insert, which was cloned into the pcDNA5/FRT/GFP construct****. The actual repetitive region is highlighted in blue.*

The loss of ATRX enrichment upon transcription at the reverse oriented telomere repeat, which generates G-rich RNA, suggests that the recruitment of ATRX is dependent on G-skew (G richness over C). The abolishment of ATRX binding is likely due to a reduction or loss of targeting signals upon inversion of the repeats.

One of the most compelling explanations for this result is related to the relationship between the possibility of forming secondary structures and the strand asymmetry of G-richness. The human SNRPN and AIRN CpG islands have previously been cloned in both orientations under the control of a constitutive viral promoter, then subjected to *de novo* methylation by DNMT3B1 and DNMT3L. When the physiological orientation was present, in which the G-rich strand is the non-template strand, the CpG islands remained unmethylated; whereas when in the opposite orientation these were permissive for DNA methylation (Ginno, Lott et al. 2012). As expected, the physiological orientation was proven to be permissive to the formation of R-loops, which protected the underlying CpG islands from interacting with the methyltransferase enyzmes (Ginno, Lott et al. 2012), providing strong evidence for strand bias in dictating events. Several other lines of evidence emphasized the importance of G-richness on the non-template strand in the efficiency of forming G4 and R-loops, two distinct structures in GC-rich DNA (Daniels and Lieber 1995, Yu, Chedin et al. 2003) (Duquette, Handa et al. 2004, Powell, Coulson et al. 2013) (Bochman, Paeschke et al. 2012). Indeed, there is a conserved tendency of G-skew on the non-template strand in many genomic high GC content regions where G-quadruplexes and R-loops have been established to have biological functions such as ribosomal DNA, Immunoglobulin Class Switch region, telomeres and CpG island promoters (Daniels and Lieber 1995, Eddy and Maizels 2008, Ginno, Lott et al. 2012), although evidence also suggests that the dependency of G4 formation on strand asymmetry of G-richness varies at the beginning and end of genes (Huppert, Bugaut et al. 2008, Fernando, Sewitz et al. 2009, Huppert 2010). G4 structures have been shown to be recognised by various proteins including numerous helicases such as human BLM and WRN (Fry and Loeb 1999), yeast Sgs1(Sun, Bennett et al. 1999),

E.coli RecQ (Wu and Maizels 2001), human XPD helicase (Gray, Vallur et al. 2014), FANCJ (London, Barber et al. 2008), RTEL1 (Uringa, Youds et al. 2011) and PIF1 (Sanders 2010). R-loops have been found to be targeted by mismatch repair factor Mutsα (Buermeyer, Deschenes et al. 1999) and by double-strand break repair factor BRCA2 (Bhatia, Barroso et al. 2014). Although no chromatin remodeling factors have been found among these, the fact that the chromatin remodeler ATRX has an ATPase/helicase domain (Argentaro, Yang et al. 2007) may give some clue about its propensity to bind DNA structures. Together with the finding in this chapter that its affinity is biased towards the transcription orientation that favors secondary structures, it is tempting to hypothesise that ATRX actually recognises such structures.

It is conceivable that ATRX recognises G-rich RNA sequences. Inverting the repeats would change G-rich transcripts to C-rich transcripts and this might explain the change in ATRX binding. Such a model is plausible because a similar phenomenon has been described in the literature. Studies using DNA FISH and RNA protected Immunoprecipitation in the Almouzni laboratory showed that sumoylated HP1α, a factor associated with the major satellite domains in pericentric heterochromatin, specifically interacted with the major forward RNA transcripts but not with the reverse ones despite both being present (Guenatri, Bailly et al. 2004) (Maison, Bailly et al. 2011). This association may serve as the seeding event in *de novo* targeting of HP1α to pericentric heterochromatin. The exact mechanism of the interaction between the protein and the forward transcripts remains to be determined; however, the binding specificity is thought to be achieved through a distinct structure that the repetitive purine rich RNA might adopt (Maison, Bailly et al. 2011). One way to

investigate this RNA-based system further would be to remove the RNA transcripts by RNase treatment before performing ATRX ChIP.

It is possible that the loss of ATRX enrichment in the reverse orientation is associated with changes in other signals such as histone modifications, histone variants or transcription factors. However, no such link between these factors and G-skew has been reported in the literature.

4.3.5 Asymmetrical distribution of ATRX across the GC-rich tandem repeats

It is intriguing that the distribution of ATRX across the physiological oriented telomere repeats is asymmetrical. If ATRX directly binds to the repeat, one would expect a peak of binding at the mid-point of the repeat with a symmetrical distribution and ATRX detectable at both ends of the repeat. In fact, an increase in ATRX binding upon transcription was undetectable upstream of the repeats, but was detectable at an increasingly high level for almost 1kb towards the 3' end of the gene (Figure 4.7A and B). This result suggests that the repeats might not directly interact with ATRX but rather trigger the binding.

This result is inexplicable by artifacts during the ChIP procedure. For example, if chromatin was under-sonicated, the high levels of ATRX at all the downstream amplicons could have resulted from Q-PCR amplification from the same large DNA fragments. However, this, if true, would fail to explain why no similar level of enrichment was detected at the upstream amplicons which are only 0.5kb away from the first downstream amplicon D1 (Figure 4.7C). Clonal variation has been ruled out because the experiment was repeated with at least two clones of each length.

One biological interpretation of this result could be that some protein were competing with ATRX for the upstream binding sites, such as RNA polymerase II that might itself be unevenly distributed across the gene possibly as a result of pausing or stalling associated with the repeats. Such a roadblock could be due to G4 structures, R-loops, G-loops (combination of G4 and R-loops) or hairpins, and is associated with an accumulation of short and abortive transcripts (Tornaletti, Park-Snyder et al. 2008, Salinas-Rios, Belotserkovskii et al. 2011, Groh, Lufino et al. 2014, Haeusler, Donnelly et al. 2014). An attempt to investigate this possibility will be described in the following chapter.

El Hage *et al*. proposed a model for the role of R-loops in blocking transcription of ribosomal DNA in a Topoisomerase1 mutant yeast strain, in which formation of R-loops was promoted by negative supercoiling behind the RNA polII. The stability of R-loops, in turn, creates a pileup of stalling RNA polII behind the loops and an asymmetrical distribution of DNA supercoiling, with positive supercoiling being accumulated in front of the loops (El Hage, French et al. 2010). A similar scenario could happen at the highly expressed ectopic gene in my experimental model and would predict that ATRX is associated with positive supercoiling or with related proteins that resolve it.

5 Exploring targeting mechanisms of ATRX

5.1 INTRODUCTION

The work described in chapter 4 aimed to investigate ATRX recruitment to ectopic GC-rich tandem repeats at a transcription inducible locus. Chromatin immunoprecipitation with an ATRX antibody was used to determine ATRX enrichment and through this to dissect the contribution of various factors including transcriptional activity, repeat size and the orientation of the G-rich strand to the localisation of ATRX. It was demonstrated that ATRX recruitment to the repeats is dependent on transcription, and in an orientation dependent manner, with ATRX recruitment favoured when the non-template strand is G-rich. Moreover, the level of enrichment is directly proportional to the size of the repeats. These findings, together with previous observations that GC-rich repeats are prone to form secondary structures, raise the possibility that a non-B form DNA structure might be a signal for ATRX targeting. However, other factors such as, histone modifications, that also alter upon transcription activation, might also play a role in localising ATRX to the GC-rich repeats.

The work in this chapter aims to address the question of what determines ATRX targeting. Such a factor should account for the observed data in the previous chapter. The main approach to the question will be to look for changes in the factor of interest when transcription is activated in the ectopic repeats and see if this correlates with the changes in ATRX enrichment. Then the factor will be perturbed and it will be seen whether this alters ATRX recruitment.

First, the histone modification H3K9me3 will be considered as it has been shown be required for the localisation of ATRX to pericentric heterochromatin directly, via interaction with the ADD domain of ATRX, and indirectly, via interaction with HP1 (Iwase, Xiang et al. 2011) (Eustermann, Yang et al. 2011). RNA polII recruits many factors during the transcription process. One thought to emerge from the observation that the ATRX distribution across the repeats seems to be cotranscriptional is that PolII may facilitate recruitment of ATRX. To investigate this further, the distribution of RNA polII will be examined. The next candidate is R-loops since these structures are formed cotranscriptionally and have directionality towards the G-rich non-template strand (see review (Skourti-Stathaki and Proudfoot 2014)). Interrogation of the role of G-quadruplexes and supercoiling of DNA will also be attempted.

5.2 RESULTS

5.2.1 Is ATRX recruited to the ectopic GC-rich repeats by Histone H3 lysine 9 trimethylation?

To interrogate the role of trimethylated lysine 9 (H3K9me3) in the recruitment of ATRX to the ectopic repeats, ChIP was carried out with an antibody specifically against H3K9me3 (Abcam) in clones containing 71 units of telomere repeats in the physiological orientation (TTAGGG)71. The beta-globin promoter and GAPDH promoter were empirically determined as a positive and negative control, respectively, in our lab. In the previous chapter, it was shown that transcription increased ATRX binding at the ectopic telomere repeat in the physiological orientation. In this chapter, I wanted to determine if there was an associated change in H3K9me3 at the repeat upon transcription. ChIP showed that there was no significant difference in this histone modification at the repeats in the induced and uninduced states (Figure 5.1), suggesting that H3K9me3 is not required for ATRX recruitment in

this case. This experiment, however, should be repeated with correction for H3 histone occupancy.

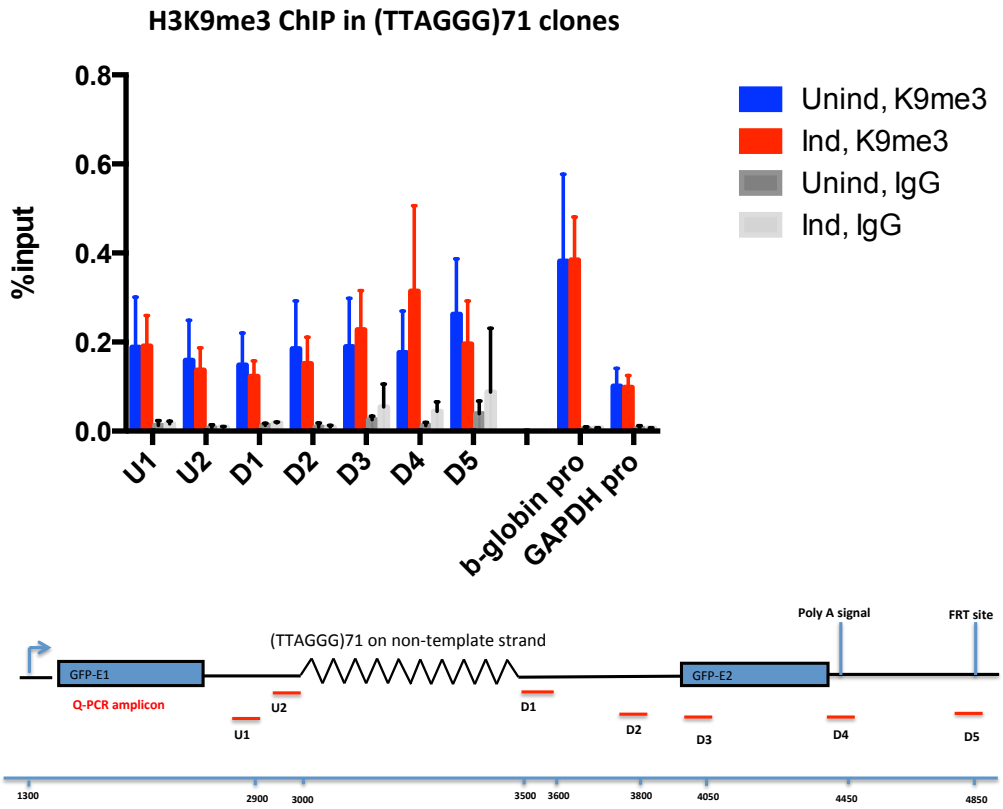

Figure 5.1 ***H3K9me3 ChIP in clones containing the ectopic 71 unit telomere repeat in physiological orientation (TTAGGG)71 when transcription is on and off (Ind and Unind, respectively).*** *Enrichment at the ectopic repeat was measured by Q-PCR amplicons indicated in the diagram and at the positive control beta-globin promoter and at the negative control GAPDH promoter. Enrichment level is expressed as % DNA input. Error bars show standard error of the mean of three independent experiments on two clones.*

5.2.2 Distribution of RNA polymerase II at the ectopic GC-rich repeats during transcription

To determine whether the distribution of ATRX depends directly on RNA polII, I performed ChIP with an antibody specific for RNA polII (N20, Santa Cruz) in the (TTAGGG)71 clones when transcription is active and inactive (Figure 5.2). The

antibody used recognises the N-terminus of the largest subunit of the enzyme, so it precipitates both promoter-bound and elongating forms of RNA polII. ChIP results show that there is an increase in PolII levels upstream and downstream of the repeats when transcription is induced, which differs from the distribution of ATRX. Surprisingly, no difference in PolII binding is observed at the promoter. Moreover, the level of PolII upstream and downstream of the repeat region is greater than at the promoter and at exon 1, which is a non GC-rich region.

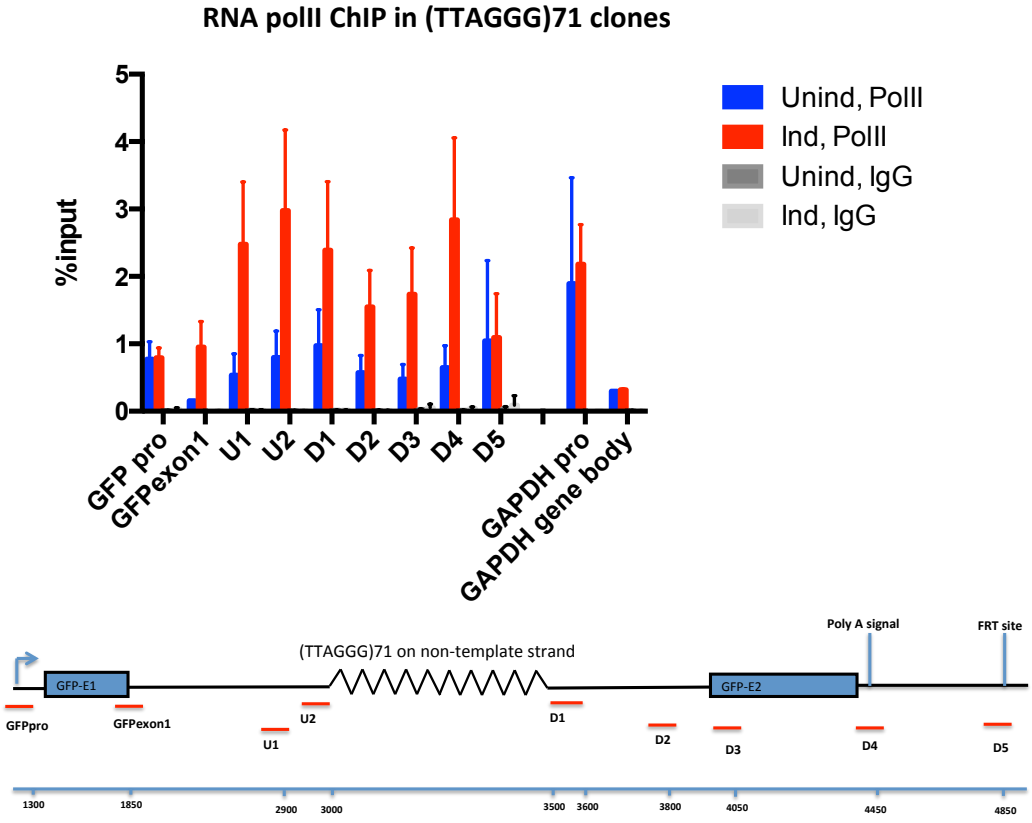

Figure 5.2 ***RNA polymerase II ChIP in clones containing the ectopic 71 unit telomere repeat in physiological orientation (TTAGGG)71 when transcription is on and off (Ind and Unind, respectively)****. Enrichment along the GFP reporter gene bearing the ectopic repeat was measured by Q-PCR amplicons indicated in the diagram, at the positive control GAPDH promoter and at the negative control GAPDH gene body. Enrichment level is expressed as % DNA input. Error bars show standard error of the mean of three independent experiments on two clones. Bar*

Unind PolII GFPexon1 and Unind PolII GAPDH gene body is based on data from one experiment.

The pattern of PolII occupancy along the active ectopic gene is, therefore, quite different from that of a typical active gene where PolII is mainly detectable at the promoter. Examples of such genes include GAPDH and ACTB (beta-actin), which are robustly transcribed housekeeping genes (ENCODE data, Figure 5.3).

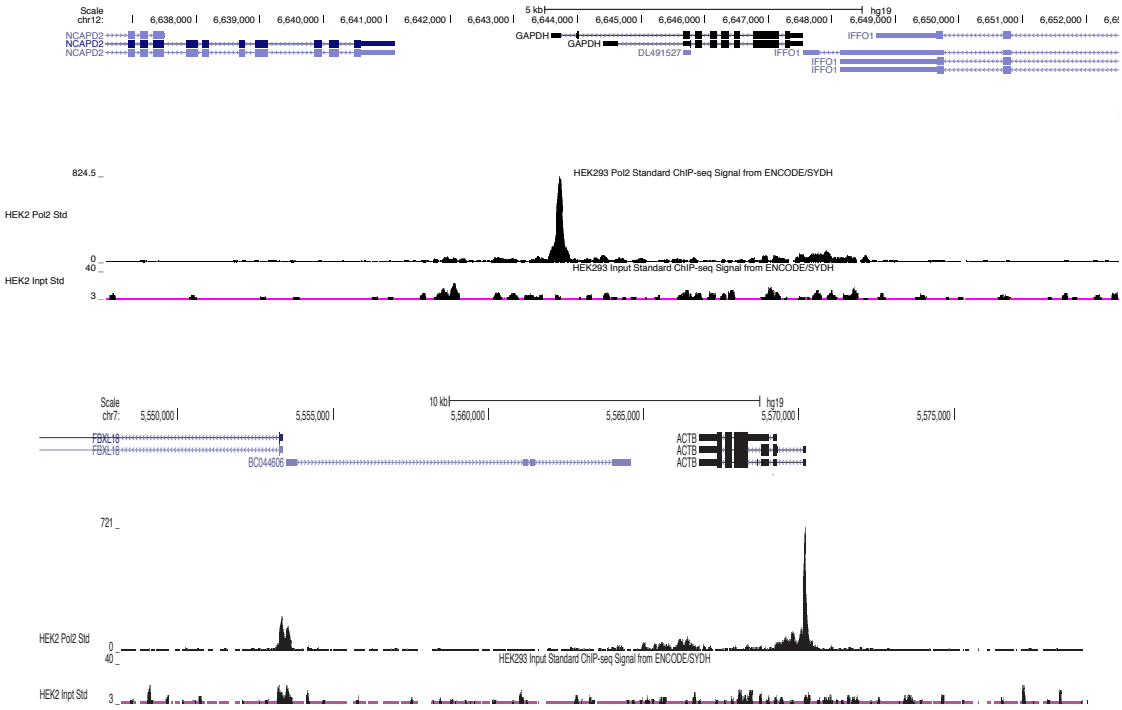

Figure 5.3 ***ENCODE data track of PolII ChIP-seq at GAPDH (upper panel) and ACTB (lower panel)****. Genes are represented by a line with exons (dark or blue bars) and introns (lines with arrows). The direction of the arrows represents the orientation of transcription.*

My PolII ChIP data on the GAPDH gene is consistent with this observation (Figure 5.2). This raises the possibility that PolII may be pausing when it encounters the repeats, which could be a consequence of a structural block.

The level of PolII around the polyA signal region (represented by the D4 amplicon) seems to be higher than that at the promoter (the GFP pro amplicon) and non-repetitive region (the GFPexon1 amplicon) of the gene, but it drops to the background level (in the uninduced state) at 400bp downstream of the polyA signal (the D5 amplicon) (Figure 5.2). This suggests that RNA polII might be slowing down at the polyA signal region immediately prior to transcription termination.

5.2.3 Does R-loop formation recruit ATRX at the ectopic GC-rich repeats?

5.2.3.1 R-loop formation at the ectopic telomere repeat in the physiological orientation

To access the occurrence of R-loops at the ectopic telomere repeat, DNA:RNA immunoprecipitation (DIP) was performed using the monoclonal antibody S9.6 which specifically recognises RNA:DNA hybrids. To control for the specificity of the antibody, the same amount of genomic DNA was treated with RNase H, which cleaves the RNA strand in the hybrid, prior to the immunoprecipitation step. Enrichment of R-loops was measured by Q-PCR and expressed as relative to GFP exon 1 in the uninduced condition. To ensure the interpretation of the experiment was correct a number of controls were used; (1) the ribosomal DNA is known to have a high level of R-loops, so was used as a positive control (El Hage, French et al. 2010) (Ginno, Lott et al. 2012). (2) DIST, a non-repetitive, non-GC rich and non-ATRX target, which is predicted not to form R-loops and (3) the Znf180 repeat, which is a known ATRX target (Hsiao Voon, personal communication) and is transcriptionally active in HEK293 cells (ENCODE data). The repeat located in Znf180 gene with a consensus size of 84bp(AGGGTTTCTCTCCAGTATGAGTTCTCTGATGTGCAACAAGTGCATAA

CTCTGGCTAAAGGATTTCCCACATTGATTACATTCAT) and a base composition A: 25.8%, C: 21.7%, G: 19.4% and T: 33.2% that is predictably not in favor of R-loop formation. The mechanism for ATRX targeting to Znf180 has not been elucidated.

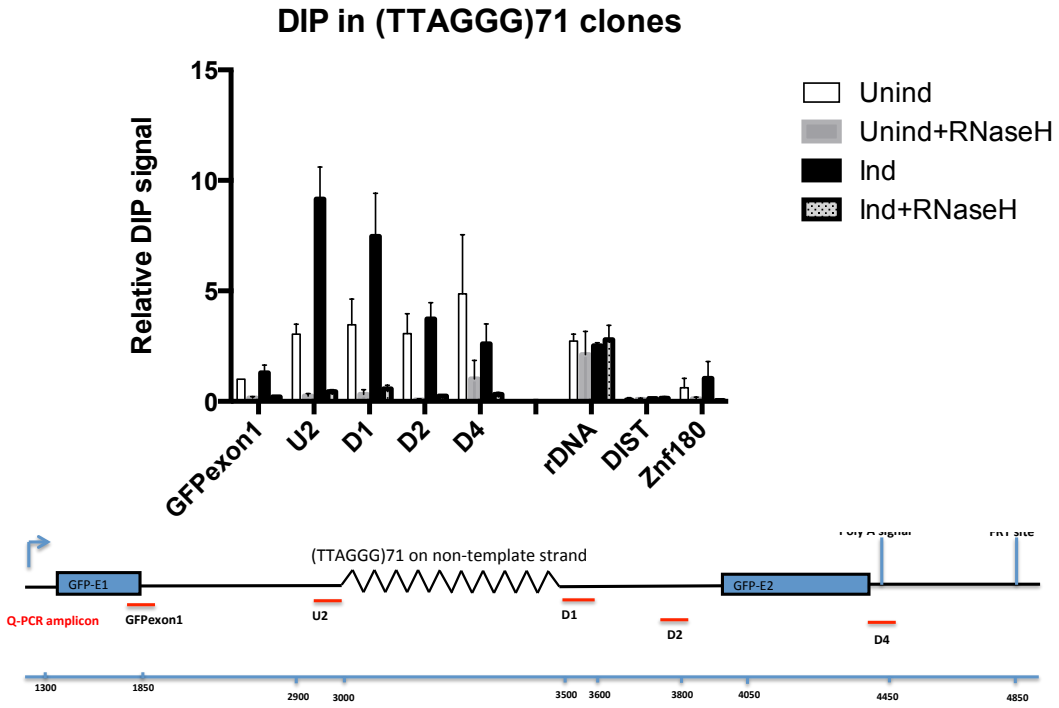

Figure 5.4 ***DIP detection in clones (TTAGGG)71****. Enrichment of R-loops at the repeat was measured by Q-PCR using amplicons indicated on the diagram, and at elsewhere in the genome. Ribosomal DNA is used as a positive control and DIST (none repetitive) as a negative non-ATRX target control. Znf180 is an ATRX target negative control. Enrichment level is normalised to DNA input and is expressed relative to GFP exon1 in the uninduced condition. Error bars show standard error of the mean of two independent experiments.*

In the physiological orientation, where the non-template strand is G-rich, 3 fold enrichment of R-loops was observed over the upstream region of the repeat (amplicon U2) in the dox-induced cells, compared to the uninduced cells. Similarly, R-loop levels increased 2 fold at the region immediately adjacent to the repeat (amplicon

D1), but not further (amplicon D2 and D4) (Figure 5.4). The specificity of the DIP signal was confirmed by RNase H treatment prior to immunoprecipitation. Following RNase H digestion, the signal was strongly reduced at all the regions across the ectopic repeat and the negative controls (Figure 5.4). However, RNase H treatment failed to remove R-loops at the positive control ribosomal DNA.

Over the course of this experiment, it was realised that RNase H did not efficiently cleave at the ribosomal RNA, and removed at most 50-60% of R-loops at this region. The source of this variation was unclear, nevertheless, it should not undermine the conclusion that genuine R-loops are formed over the ectopic telomere repeat of the physiological orientation upon induction of transcription.

It is worth noting that in the uninduced cells, R-loop signals over the regions U2 and D1-D4 upstream and downstream of the repeats are comparable to that at the positive control rDNA and higher than the background at GFP exon1 and the negative control DIST and Znf180 (Figure 5.4), indicating that R-loops may also form in these cells. This could be due to leaky and/or basal transcription and the high stability of R-loops.

5.2.3.1.1 R-loop formation at the ectopic ψζVNTR in the physiological orientation

Next, to examine R-loop formation at the ectopic ψζVNTR in the physiological orientation, DIP was carried out in a clone containing 390bp (GCGGG) repeat. Preliminary data have shown that R-loops form over the ψζVNTR upon transcription induction (Figure 5.5), with a biased distribution towards the upstream region of the repeat (amplicon U1) compared to the downstream regions (amplicon D1, D2 and D4). The experiment is to be repeated to confirm this result.

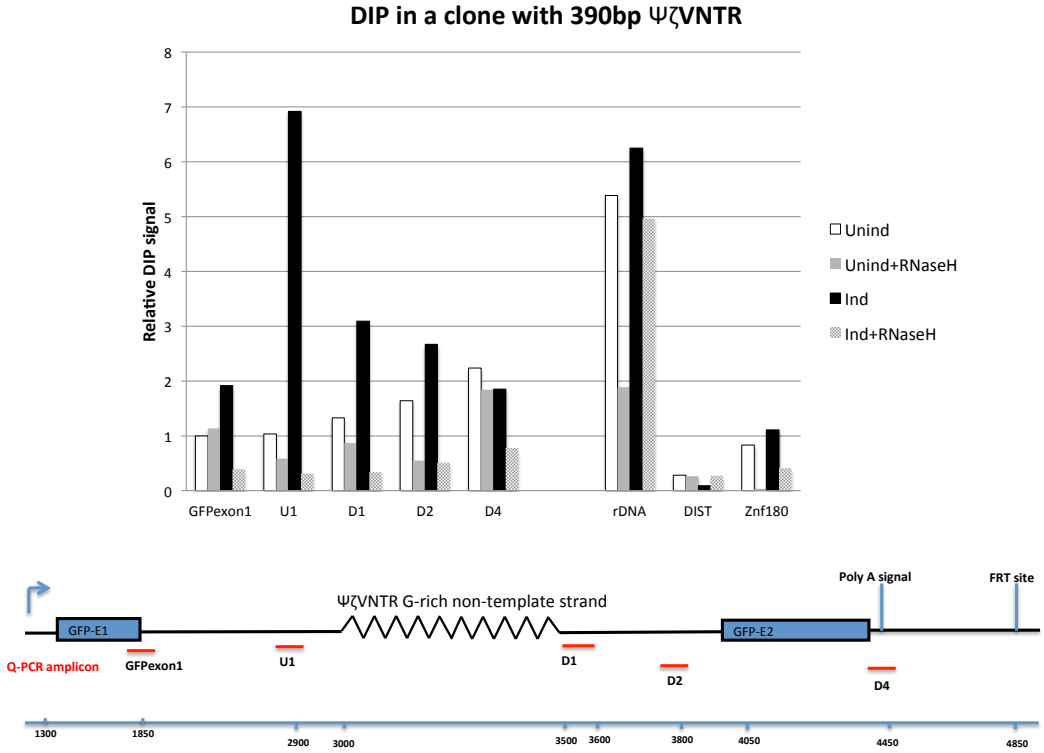

Figure 5.5 ***DIP in a clone containing a 390bp ΨζVNTR of the physiological orientation.***

5.2.3.1.2 R-loop formation at the ectopic telomere repeat in the reverse orientation

Using *in vitro* and episomal transcription systems, previous studies show that R-loops preferentially form on plasmids when the non-template (non-transcribed) strand is G-rich (Yu, Chedin et al. 2003, Li and Manley 2005, Ginno, Lott et al. 2012, Powell, Coulson et al. 2013). Similarly, genome-wide analyses of CpG island promoters, where R-loops have been demonstrated to occur, show a G-rich bias on the non-template strand downstream of the transcription start sites (Ginno, Lott et al. 2012). To test if this is the case *in vivo* and on chromatin, DIP was performed in the clones containing 71 units of ectopic telomere repeat in the reverse orientation (CCCAAT)71 where the G-rich strand is transcribed.

Previously, I have shown that ATRX recruitment increased upon transcription of the ectopic telomere repeats in the physiological orientation, which favours R-loop formation, but not in the reverse orientation. DIP analysis in the reverse-oriented telomere repeats will be done here to address the important question whether there is a correlation between R-loop formation and ATRX recruitment.

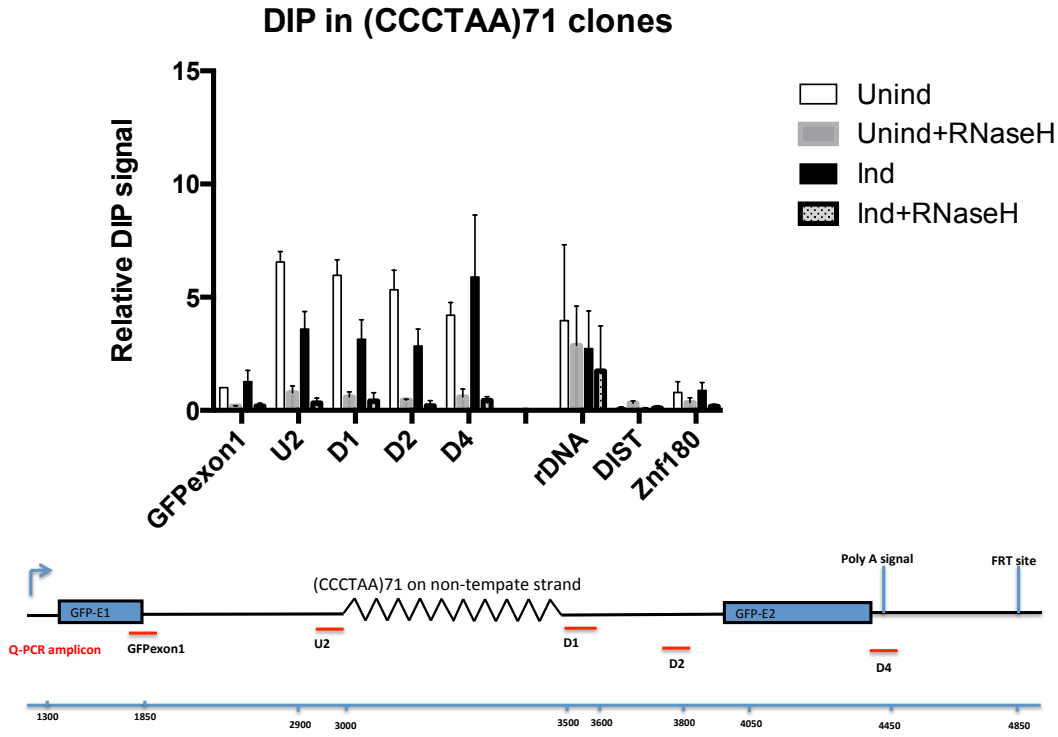

Figure 5.6 ***DIP detection in clones carrying the ectopic 71 unit telomere repeat of the reverse orientation (CCCTAA)71****. Enrichment of R-loops at the repeat was measured by Q-PCR using amplicons indicated on the diagram, and at elsewhere in the genome. Ribosomal DNA is used as a positive control and DIST (none repetitive) as a negative non-ATRX target control. Znf180 is an ATRX target negative control. Enrichment level is normalised to input DNA and is expressed relative to GFP exon1 in the uninduced condition. Error bars show standard error of the mean of three independent experiments (Bar Unind and Ind), and of two independent experiments (Bar Unind+RNaseH and Ind+RNaseH).*

Surprisingly, DIP in the reverse-oriented telomere repeat clones detected R-loops at the repeats before and after transcription activation, in comparison to the negative and positive controls (Figure 5.6). The DIP signals were verified by RNase H treatment of

the genomic DNA prior to immunoprecipitation with the S9.6 antibody. Oddly, the transcription-uninduced condition (No Dox) produced approximately 2 times more R-loops than the transcription-induced condition (plus Dox) (Figure 5.6). The reason for this is unclear but one possible explanation is that there is antisense transcription from an unknown endogenous promoter, and that is suppressed by sense transcription from the ectopic promoter.

5.2.3.2 How does R-loop destablisation affect ATRX binding at the ectopic repeats?

Next, to attempt to determine definitively if R-loops trigger ATRX recruitment I performed ATRX ChIP in Rnase H1 overexpressing clones carrying the physiological orientation of ectopic telomere repeats. To overexpress RNase H1, a vector containing the RNase H1-GFP fusion gene (Skourti-Stathaki, Proudfoot et al. 2011) (kindly provided by Nicholas Proudfoot's laboratory) was transiently transfected into the clones. Transfection into the 293T-REx cells was optimised using a series of Lipofectamine 2000 concentrations and the maximal efficiency of 42-45% was achieved with 15µL Lipofectamine 2000 (Figure 5.7), as measured by FACS.

Following 24 hours of RNase H1 overexpression, the cells were incubated with 1µg/mL doxycycline containing medium for another 24 hours to activate transcription of the ectopic telomere repeat before being harvested for ChIP of ATRX. ChIP in two clones shows no significant difference in ATRX recruitment to the ectopic telomere repeats in cells overexpressing RNase H1, compared to the wild-type cells (Figure 5.8).

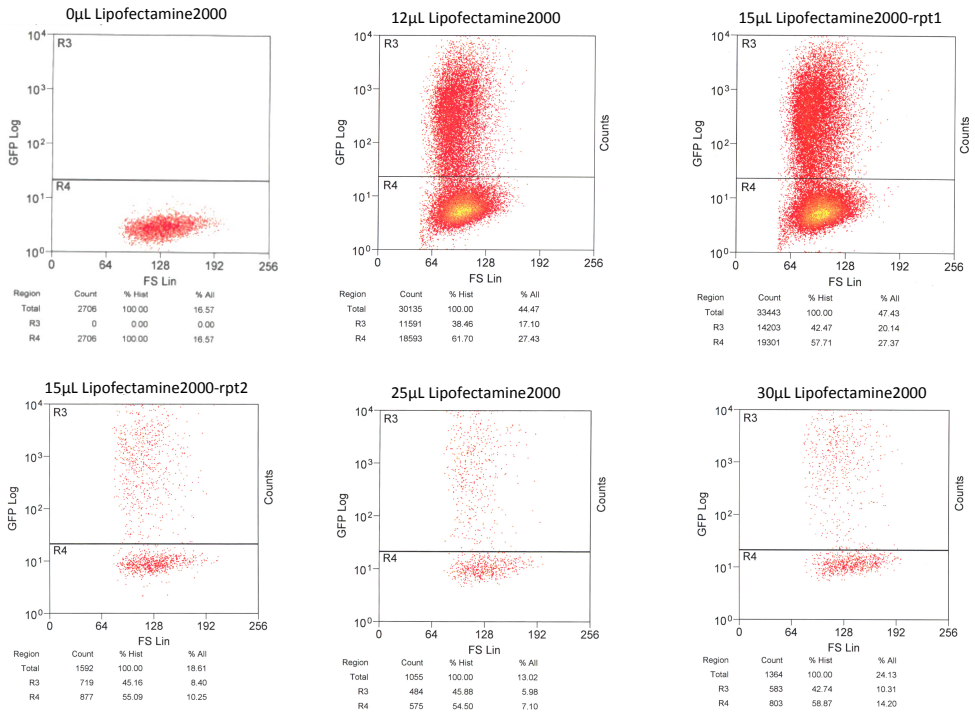

Figure 5.7 ***Optimisation of transient transfection of RNase H1–GFP fusion protein expressing plasmid into the stable expression clones****. A series of indicated amount of Lipofectamine 2000 was used to transfect Rnase H-GFP fusion expression plasmid into 293T-REx clones. After 24 hours of transfection and incubation in No Dox condition, the cells were harvested and efficiency of the transfection was checked by the percentage of cells expressing GFP using FACS analysis.*

There are three possible explanations for this result. First, transfection efficiency was not sufficient to degrade R-loops at the ectopic repeats. Secondly, the overexpressed RNase H failed to resolve R-loops. Lastly, it is possible that perturbation of R-loops did not cause any change in ATRX binding at the ectopic repeat, indicating that they are not *bona fide* triggers of ATRX recruitment in this case. However, this conclusion can only be confirmed by repeating the experiment with more careful quantitation of recombinant RNase H levels and verification of efficient R-loop removal by DIP in transfected cells.

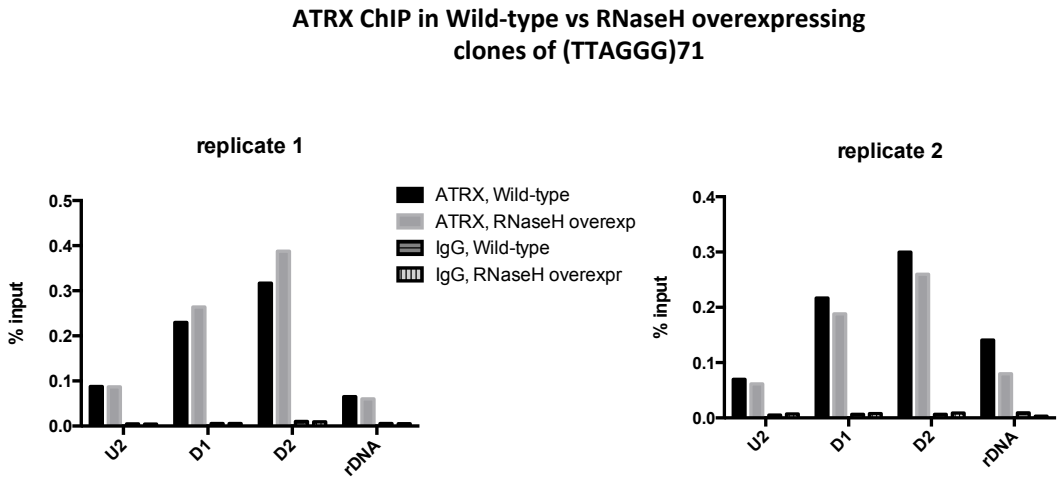

Figure 5.8 ***ATRX ChIP in Rnase H1 overexpressing clones****. RNase H1-GFP expression plasmids were transfected into two clones (TTAGGG)71 Tel71(6) and Tel71(7). After 24 hours of transfection, doxycycline (1μg/mL) was added into culture to induce transcription of the repeats. The cells were incubated for another 24 hours before being harvested for ATRX ChIP. The experiment were repeated in two different (TTAGGG)71 clones.*

5.2.4 Stablisation of G-quadruplexes by Pyridostatin reduces ATRX binding

Telomere repeats are known to form G-quadruplexes. This process is facilitated by DNA replication and transcription during which the DNA duplex is transiently denatured into single strands. To investigate how stabilising G-quadruplex structure affects ATRX binding, ChIP was conducted with the ATRX antibody H300 in the (TTAGGG)71 clones in the presence and absence of a G4 stabilising compound, called Pyridostatin.

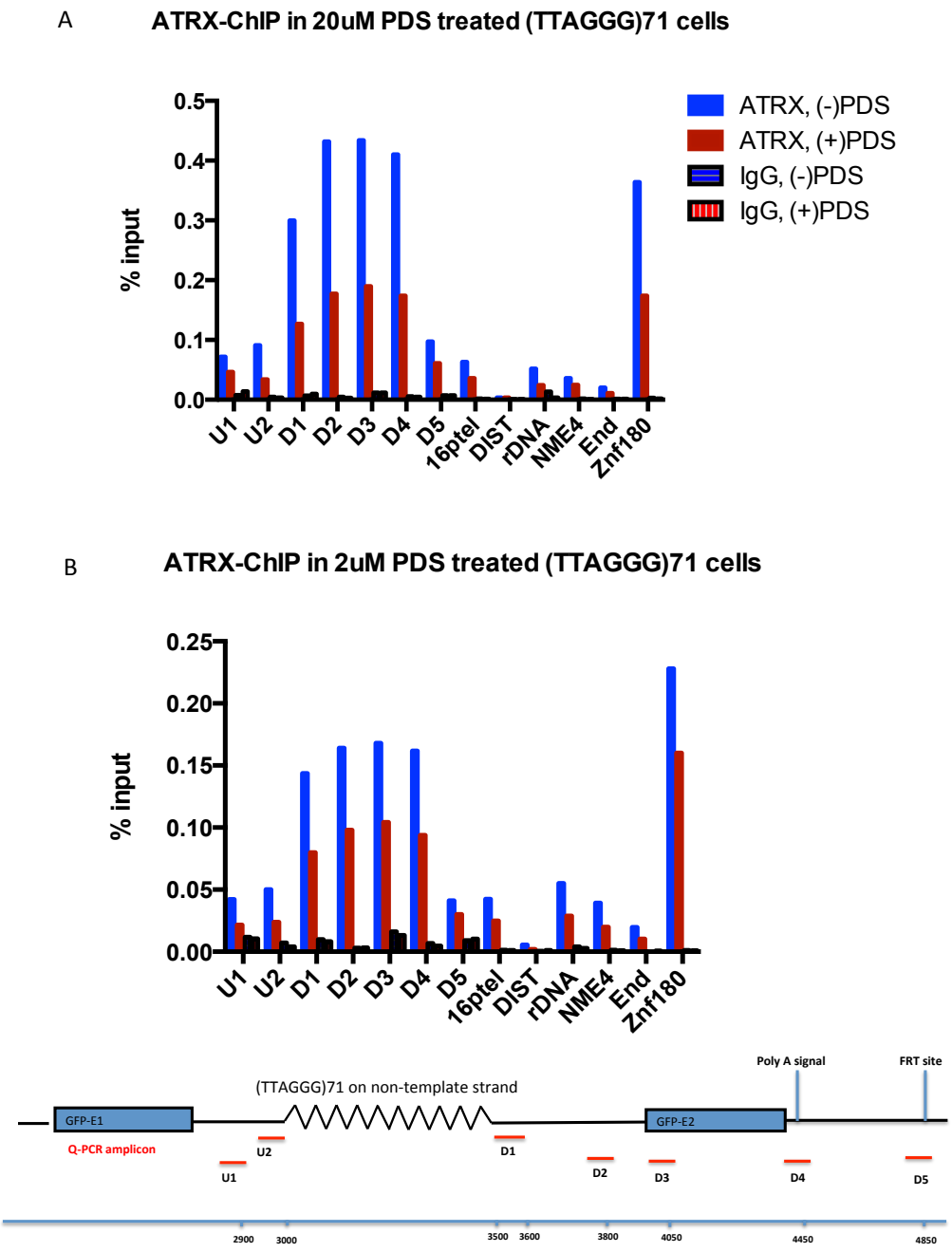

Figure 5.9 ***ATRX ChIP in PDS treated cells****. A (TTAGGG)71 clone was treated with (A) 20μM Pyridostatin (PDS) 4 hours or (B) 2μM PDS 24 hours prior to transcription induction by doxycycline. Cells were harvested 24 hours later for ChIP with ATRX antibody. ATRX enrichment was measured by Q-PCR.*

It was previously reported that Pyridostatin can specifically bind and stabilise G-quadruplexes, causing genome instability and modulation of gene regulation

(Rodriguez, Muller et al. 2008, Muller, Sanders et al. 2012, Rodriguez, Miller et al. 2012, Lam, Beraldi et al. 2013).

The (TTAGGG)71 clones were treated with 20μM PDS overnight or 2μM PDS for 24 hours, followed by induction of transcription of the ectopic repeat by the addition of 1μg/mL doxycycline. ATRX ChIP was then carried out on the treated cells and untreated cells (without PDS and with doxycycline addition). It is predicted that if ATRX binds G-quadruplexes, the stabilisation of this structure would lead to an increase in ATRX enrichment level at the ectopic repeat and other G4 motif DNA. Surprisingly, PDS treatment led to a 2-3 fold reduction in ATRX binding at the ectopic repeat at 20μM concentration and 1.6-2 fold at 2μM concentration, compared to the PDS-untreated control cells (Figure 5.9A and B). In addition, ATRX enrichment levels at other G4-rich sequences (rDNA, 16p telomere, NME4 and endogenous ψζVNTR) were also decreased (Figure 5.9A and B). This could be due to PDS blocking or ablating binding sites of ATRX. Recent studies show that TMPyP4, a cationic porphyrin G4 ligand, distorts G-quadruplexes formed in C9orf72 and MT3-MMP mRNA and disrupts RNA-binding protein interactions with the G4 motif in C9orf72 (Zamiri, Reddy et al. 2014) (Morris, Wingate et al. 2012). However, at both concentrations of PDS, a similar reduction of ATRX signal was also observed at the repeat Znf180 control (Figure 5.9), which is an ATRX target (Hsiao Voon, personal communication) and a non-G4 sequence as tested by **Q**uadruplex forming **G**-**R**ich **S**equences (**QGRS**) mapper software (Kikin, D'Antonio et al. 2006). This suggests that PDS might work indirectly through down-regulating ATRX, or affecting ATRX binding, which inhibits its interactions with chromatin.

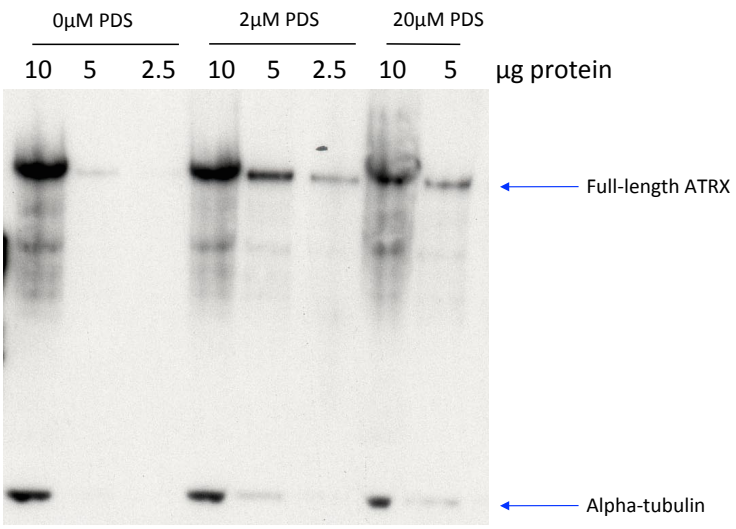

Figure 5.10 ***Western Blot analysis of ATRX protein levels in the PDS treated and untreated cells****. The (TTAGGG)71 clone was cultured in the medium containing no PDS, 2μM or 20μM PDS for 24 hours. After that, protein was extracted and used for the analysis. 10, 5 or 2.5μg of total protein of each concentration of PDS was run and probed with ATRX antibody H300. Alpha-tubulin is used as the loading control.*

To check whether PDS affects ATRX expression, a Western Blot analysis was performed to compare the level of ATRX in the untreated, 2 and 20μM PDS treated cells. Figure 5.10 shows that there is no obvious change in the ATRX protein levels in the treated compared to the untreated cells, indicating that PDS does not affect ATRX expression. It remains for further investigation to as whether PDS might block ATRX binding.

5.2.5 Does supercoiling trigger ATRX binding at the ectopic repeats?

The fact that the ATRX distribution over the ectopic telomere repeat region is asymmetrical and yet the R-loop level is symmetrical at both 5' and 3' ends of the region hints that ATRX may not directly associate with R-loops. However, the loops may indirectly stimulate ATRX binding by causing positive supercoiling to accumulate in front of the loops.

Positive supercoiling is overwound DNA that can be formed ahead of RNA polII during transcription (Wang and Lynch 1993, Cook 1999, El Hage, French et al. 2010). It becomes problematic if unresolved by helicases such as Topoisomerase I (Top I) and Top II (for review, see(Wang 2002)). In the absence of Top I and RNase H, persistent R-loops could lead to a high level of stable positive torsion in front of the loops and simultaneously stabilise negative torsion behind the loops (El Hage, French et al. 2010). This scenario could happen at a highly transcribed gene like the ectopic GFP gene and if true would predict an association of ATRX with positive supercoiling.

To test this hypothesis, I employed the use of Bleomycin, a glycopeptide antibotic exclusively used in cancer therapy that is able to rapidly release torsional stress by introducing DNA strand nicks (Villeponteau and Martinson 1987, Naughton, Avlonitis et al. 2013). The (TTAGGG)71 clones were incubated with doxycycline for 24 hours to induce transcription of the ectopic locus. Before harvesting the cells, 100μM Bleomycin was added to the culture and incubated for 10 minutes. To see if ATRX associates with positive supercoiling, ATRX ChIP was performed in Bleomycin treated and untreated cells.

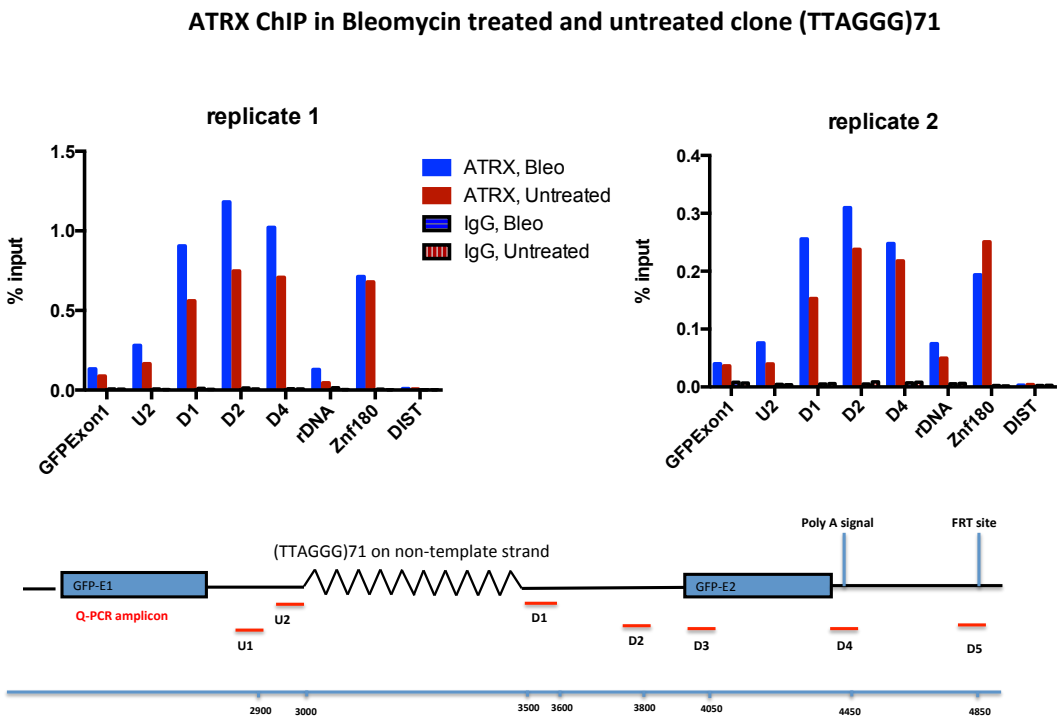

Figure 5.11 ***ATRX ChIP in the Bleomycin treated cells****. ChIP was performed for cells treated with 100μM Bleomycin for 10 minutes, following transcription induction by doxycycline. Replicate 1 and 2 are two different (TTAGGG)71 clones.*

Surprisingly, ATRX binding levels at the ectopic repeat region in the Bleomycin treated cells went up by 1.5 – 2 fold compared with the untreated cells (Figure 5.11), indicating that ATRX may not directly bind positive supercoiled DNA. However, it is important to note that ATRX localises to DNA damage sites (Leung, Ghosal et al. 2013), which could explain why ATRX was recruited more to the repeat as potential DNA damages might result from the Bleomycin treatment at this region. To investigate this further, Top I and Top II inhibitors can be used, which prevent supercoiling relaxation without causing DNA breaks.

5.3 DISCUSSION

5.3.1 Histone 3 lysine 9 trimethylation is not required for recruitment of ATRX to the ectopic tandem repeats

Lysine 9 trimethylation (K9me3) is a well-known heterochromatin repressive mark, and is highly enriched at telomeres and centromeres. However, there are a subset of actively transcribed genes that also harbour K9me3 in their bodies (Hahn, Wu et al. 2011). In the model system used here, this histone mark is enriched at a low level in the body of the ectopic gene GFP (Figure 5.1).

K9me3 is required for the targeting of ATRX to pericentric heterochromatin, as well as contributing to its enrichment at a group of low GC content repeats. When Suv39H1 and Suv39H2 methyltransferases that are responsible for adding methyl groups are depleted, ATRX localisation at pericentric heterochromatin is completely abrogated (Kourmouli, Sun et al. 2005, Eustermann, Yang et al. 2011). Structural analysis of the ADD domain of ATRX revealed that it directly binds K9me3, playing a central role in the recruitment mechanism of ATRX to this region. Furthermore, K9me3 also recruits ATRX via interacting with HP1 (Eustermann, Yang et al. 2011) (Iwase, Xiang et al. 2011). However, a Suv39H1 and Suv39H2 double knock-out does not reduce, but rather leads to an increase in the telomeric localisation of ATRX. Interestingly, a similar study by Voon *et al*., shows that the ATRX localising to GC-rich DNA appears to be independent of H3K9me3 (Hsiao Voon, personal communication).

Furthermore, a recent study by Ivanauskiene *et al*. suggests that targeting of ATRX to telomeres in embryonic stem cells is guided by an oncoprotein, called DEK, loss of which leads to impairment of ATRX recruitment at telomeres (Ivanauskiene, Delbarre

et al. 2014). A message taken from these studies is that targeting of ATRX to different genomic locations seems to require different signalling mechanisms. The ectopic telomere repeat described in this study represents high G-rich repetitive DNA in active regions of the genome, and maybe requires a different signal rather than K9me3 for directing ATRX. Indeed, there was no change in the level of this modification that could account for the increase in ATRX enrichment at the ectopic repeat region during transcription (Figure 5.1). The data in this experiment could have been strengthened with a correction for histone H3 deposition.

5.3.2 RNA polII might be accumulating at the repeats

Active genes are typically characterised by an abundance of RNA polII occupancy at transcription start sites compared to intragenic regions (Sun, Wu et al. 2011). The capability of ChIP to capture a protein at a specific region relies on how dynamic it is at that region. Accordingly, at promoters/transcription start sites of active genes, where RNA polII is accumulated/poised to be release into the gene body, it can easily be caught by fixation for ChIP. In contrast, as the enzyme moves quite fast during elongation it is hard to be detected by ChIP. RNA polII enrichment at GAPDH promoter and gene body in the clones is consistent with this notion. It is also supported by examples of several genes from ENCODE data of RNA polII ChIP-seq (Hek 293 pol2 ENCODE/SYDH data).

Although gene activation is usually achieved via an increase in PolII at the promoter, in particular genes it is mediated by a change in frequency of releasing PolII without the need for more recruitment (Mokry, Hatzis et al. 2012). Moreover, the rate of PolII release from the promoter and the rate of elongation affect the ability of PolII detection by ChIP. Based on these observations, it is possible that upon activation by

doxycline, PolII was recruited but released so quickly from the GFP promoter that any change in its enrichment compared to the uninduced state was not detected by ChIP (Figure 5.2).

Furthermore, it has been shown that the rate of elongation by RNA polII is associated with certain gene features. For example, long repeats and high GC content were found to be negatively correlated with elongation rate in several cell lines (Veloso, Kirkconnell et al. 2014). The low level of PolII at the end of exon 1 compared to upstream and downstream of the repeats (shown by the U1 to D4 amplicon) in the GFP gene indicates that RNA polII might be progressing fast through the non-repetitive region (of exon 1) but slowed down when it "hit" the repeat (Figure 5.2). ChIP using an antibody that recognises elongating PolII would clarify this further.

The fact that PolII levels stayed high downstream of the repeat and towards the end of the gene suggests that the enzymes did not completely stall at the 5' end of the repeat (U1 and U2 bar, Figure 5.2), but might just transiently pause and restart.

Movement of RNA polymerase II might be impeded if there is a structural block during the process of transcription, and would provide evidence of the existence of such a block. In addition, one compelling possibility is that accumulated RNA PolII might compete for the binding site of ATRX, and this might explain the asymmetrical distribution of ATRX upon transcription activation described in chapter 4. However, the distribution of PolII observed does not seem to reflect this.

5.3.3 Correlation of R-loop and G-quadruplex formation with ATRX recruitment at the ectopic GC-rich repeats

The DNA immunoprecipitation and ATRX ChIP results suggest that the presence of R-loops correlates with the recruitment of ATRX to the ectopic telomere repeat. R-loops are structures formed co-transcriptionally at high GC content sequences. Moreover, these structures were shown to preferentially form when the non-template strand is G-rich. In my model system R-loop formation and ATRX binding are concomitantly enriched upon transcription of the ectopic telomere repeat in the physiological orientation with high G content on the non-template strand. Furthermore, the transcription-dependent increase in ATRX enrichment at the downstream region of the repeat is not seen in the reverse orientated repeat, which does not favour R-loop adoption.

Interestingly, the DIP results also show that at the reverse oriented repeat the R-loop level in the transcription-off condition is higher than that in the transcription-on condition (Figure 5.6). This does not argue against the correlation between R-loop occurrence and ATRX recruitment. In fact, it looks as though the ATRX level at the upstream region (bar U1, Figure 4.8B) of the un-transcribed repeat is higher than the transcribed repeat, supporting the correlation. An obvious question is how do more R-loops form when the gene is not transcribed compared to when it is transcribed, given an unarguable fact that R-loops are co-transcriptional structures. One scenario would be the presence of antisense transcription initiated from an external promoter that is located downstream of the ectopic construct locus. This antisense transcription might be impeded by the robust sense transcription, when this is induced in the presence of doxycycline. This hypothesis is justified for two reasons. First, it complies with the co-transcriptional formation rule. Second, in the antisense transcription, the G-rich

template strand of the sense orientation becomes the non-template strand, which favours R-loop formation. In support of this hypothesis, ATRX distribution at the reverse oriented repeat in the un-induced condition is asymmetrical with more binding at the 3' end of the reverse repeat (bar U1, Figure 4.8B) in relation to antisense transcription direction, which mirrors the distribution of ATRX over the physiological oriented repeat in the sense transcription. More Q-PCR analysis using amplicons further upstream (towards the GFP exon 1) is needed to draw a more solid conclusion. To investigate the possibility of antisense transcription, Nanostring analysis can be performed to look directly at the sense and antisense RNA transcripts.

Attempts to dissect the role of R-loops in triggering ATRX recruitment by overexpressing RNase H to perturb loop stabilisation have not been completed since DIP analysis to verify R-loop removal by the overexpressed enzyme was not done due to time constraints. In addition, stabilising R-loops by knocking down the endogenous RNase H can also be done to see if ATRX binding is increased further at the repeat. This will complement the RNase H overexpression experiment in addressing the question as to whether R-loops recruit ATRX.

Although telomere repeats are well-know to form G-quadruplexes (Parkinson, Lee et al. 2002, Biffi, Tannahill et al. 2013), it was necessary to examine if the formation of this structure increased upon transcription in my model system, which would help elucidate its role in recruiting ATRX. However, my attempts to immunoprecipitate G-quadruplexes *in vivo* using antibody 1H6 (Henderson, Wu et al. 2014) have been unsuccessful (data not shown). A further attempt to interrogate the association of G-quadruplexes and ATRX recruitment by treating the cells with the G4 stabilising drug

PDS needs other non-G4 forming control sequences. Other G4 stabilisers could also be employed, e.g telomestatin.

G-quadruplexes and R-loops can coexist in a structure called G-loops during transcription (Duquette, Handa et al. 2004) (Duquette, Pham et al. 2005). The formation of RNA:DNA hybrids, that leaves unpaired guanines in the non-template strand free, facilitates the formation of G-quartets. It is possible that G-loops are formed at the ectopic repeats in my model system and these trigger ATRX recruitment.

One question still to be answered is whether ATRX directly binds to one of these discussed secondary structures. The asymmetrical distribution of ATRX across the ectopic repeats (Figure 4.7) suggest that the chromatin remodeler may not directly interact with R-loops because the enrichment of these structures seems to be symmetrical (Figure 5.4), with a transcription-dependent increase upstream (bar U2, Figure 5.4) of the repeat where no increase in ATRX enrichment was detected. Furthermore, G4 structures, if formed within the repeats, would be limited to the G-rich sequence fragment, but ATRX binding is observed downstream of this sequence. This also argues against the hypothesis that ATRX directly interacts with G4.

The correlation between the occurrence of the R-loops and ATRX enrichment during transcription predicts that they may trigger ATRX recruitment. The formation of R-loops or other possible secondary structures at the repeat could lead to an accumulation of positive supercoiling ahead (of R-loops) or to the recruitment of an unknown factor that associates with ATRX.

6 General discussion and future perspectives

Chromatin remodelers are present at high abundance within cells, and are recruited to specific locations to remodel chromatin as part of many cellular activities. As the first step of the remodeling process, it is crucial that a remodeler is localised to the right sites. This step is entirely dependent on a specific feature of the targeted region serving as a signal for the remodeler localization. This DPhil project aimed to understand how the chromatin remodeler ATRX identifies a subset of its targets, the GC-rich tandem repeats, in the genome. In other words, what are the specific features of the repeats that determine ATRX recruitment?

6.1 ATRX is recruited to GC-rich tandem repeats independently from the genomic context

ATRX localisation at GC-rich tandem repeats has been well-documented, but whether this localisation is dependent on the chromatin context or the repeat features is unknown. Here, I demonstrate that ATRX is recruited to the GC-rich tandem repeats independently from the surrounding chromatin context. This conclusion is drawn from the observation that when taking a GC-rich tandem repeat that was defined as an ATRX target by ChIP from its endogenous environment and putting it into a random location in the genome (ectopically), ATRX binding is maintained.

6.2 ATRX recruitment to GC-rich tandem repeats independently from the H3K9me3 histone modification

H3K9me3, which recruits ATRX to pericentric heterochromatin, is not required for ATRX localisation at the ectopic GC-rich tandem repeats. These observations suggest that a different mechanism is employed to direct ATRX to these types of targets. It has been demonstrated that XNP, the fly homolog of ATRX, binds directly to nucleosome-depleted chromatin in a highly transcribed gene (Schneiderman, Orsi et al. 2012). It remains to be determined whether nucleosome occupancy also plays a role in ATRX localisation.

6.3 Evidence of a structural feature as the targeting signal for ATRX localisation

Data in this study provide evidence that ATRX recruitment to the GC-rich tandem repeats might be triggered by a structural feature adopted by the repeats. This speculation is based on three findings. First, the efficiency of ATRX binding positively correlates with the repeat size. Previous studies have suggested that the repeat size increases the propensity of sequence to adopt G-quadruplexes and R-loops. Therefore, it is tempting to speculate that the longer the repeat the more it is prone to adopt these structures, which leads to the increase in the ATRX binding.

Moreover, ATRX targeting at the ectopic GC-rich repeats is facilitated by their transcriptional activity. This finding might account for the tissue specificity of some ATRX targets, an observation made in the previous study by (Law, Lower et al. 2010). Importantly, by loosening the chromatin, separating the two strands of the

DNA duplex, and inducing negative superhelicity, transcription provides opportunities for the formation of non-canonical structures, such as G4s and R-loops. This second finding implies that ATRX could be recruited by these structures. However, it is also possible that the recruitment is promoted by transcription factors or RNA polII. A previous genome-wide study in our lab shows that most ATRX targets which are rich in GC content are transcribed (Hsiao Voon, personal communication). In future work, it would be valuable to explore this further.

Interestingly, the possibility that RNA polII and transcription factors are involved in ATRX recruitment is ruled out by the third finding that the ATRX binding is dependent on the transcriptional orientation of the G-rich strand being the non-template strand. It is unlikely that the recruitment of the transcription machinery, including RNA polII and transcription factors is affected by inverting the orientation of the G-rich strand, given the fact that the associated reporter gene GFP is still expressed in both orientations. On the other hand, G-richness on the non-template strand, rather than on the template strand, has been shown to be preferred for G4 and R-loop formation, supporting the hypothesis that ATRX actually recognises these structures.

In support of the formation of a structure at the GC-rich tandem repeats, data on the RNA polII ChIP indicate an accumulation of the enzyme over the repeat region.

6.4 Which structure triggers ATRX targeting?

The question of which structure ATRX recognises in the GC-rich tandem repeats is addressed in chapter 5. While attempts to explore the presence of G4s in the ectopic locus and the influence of stabilising this conformation on ATRX recruitment are hampered by technical difficulties, the striking asymmetrical binding of ATRX with a higher level detected downstream of the repeats weakens the hypothesis that G4s interact with ATRX. If ATRX binds G4s, one might expect the binding levels at both sides of the repeat to be similar, as the concentration of G4 motifs is the highest in the middle of the repeat region. R-loops, however, have been shown to extend beyond G-rich regions and invade into adjacent sequences (Roy, Zhang et al. 2010). It is more plausible that these structures interact with ATRX. Another possibility is that both G4s and R-loops may not directly bind ATRX but trigger a change downstream of the repeats that directly associates with the remodeler.

Future work would be informative in further investigating whether asymmetrical distribution of ATRX also occurs at other GC-rich tandem repeat targets in the genome.

Chapter 5 continues the exploration of these possibilities. The data on R-loop formation show that R-loops are indeed adopted and increased with transcription in the physiological orientation of the G-rich strand, which correlates with the increase in the ATRX binding in this orientation. This correlation is also supported by the results of R-loop enrichment in the reverse oriented repeats, although in future work, antisense transcription through the ectopic locus should be examined using RT-PCR

or more advanced methods such as NanoString, which can directly detect strand specific transcription. Together, these observations suggest that R-loops may trigger ATRX targeting to GC-rich tandem repeats.

6.5 R-loops: the direct or indirect targeting signal?

The distribution of the R-loops does not seem, however, to mirror that of ATRX binding across the repeats, suggesting that ATRX might not directly bind this hybrid conformation. Furthermore, the destabilisation of the R-loops by RNase H overexpression did not reduce ATRX binding. One caveat of this experiment is that the effectiveness of R-loop removal by the overexpressed RNase H was not assessed. Future experiments will address this. In addition, R-loop stabilisation by RNase H knock down could be performed to see if ATRX binding is increased. This would complement the RNase H overexpression experiment in interrogating the interaction of ATRX with R-loops.

The formation of R-loops at the repeats could lead to an accumulation of positive supercoiling ahead of the loops, and pausing of RNA polII could enhance this accumulation (El Hage, French et al. 2010). This scenario seems to fit with the ATRX binding bias towards downstream of the repeats. Although an attempt was made in this study to investigate the association of ATRX with positive supercoiling using Bleomycin, a compound relaxing torsional stress by creating DNA nicks, this drug could cause double-strand breaks and hence attract ATRX (Leung, Ghosal et al. 2013), which may confuse the interpretation. Better approaches with minimal side effects, such as Topoisomerase I inhibition, are needed for future studies.

Accumulation of transcription-generated positive torsion by topoisomerase inhibition destabilises nucleosomes (Teves and Henikoff 2014). Moreover, it is evident that XNP, the *Drosophila* homolog of ATRX, directly binds nucleosome-depleted chromatin of highly transcribed genes (Schneiderman, Orsi et al. 2012) where it facilitates the deposition of H3.3 histone variant. The histone chaperone HIRA has also been demonstrated to directly bind non-nucleosomal DNA as a protective nucleosome gap filling prior to H3.3 incorporation. Together these studies suggest that a second scenario for ATRX targeting to the GC-rich repeats in this study. That is, the R-loop formation accumulates positive torsional stress, which leads to nucleosome depletion downstream of the repeats, and this in turn recruits ATRX. This hypothesis can be tested by micrococcal nuclease protection assay, which is based on the fact that nucleosome-free chromatin is sensitive to this enzyme.

6.6 The significance of this study in addressing the variation in alpha-thalassemia in ATR-X syndrome

The negative correlation between the length of adjacent GC-rich tandem repeats and associated gene expression in the absence of ATRX has been documented previously (Law, Lower et al. 2010). This study demonstrates that ATRX binding positively correlates with the repeat size. The probability of forming non-canonical structures also increases with the repeat size, which attracts more ATRX. An interpretation of this finding is ATRX recognises a structured feature induced by the GC-rich tandem repeats and exerts its role on resolving this structure or stabilising the chromatin. In the absence of ATRX, the longer the repeats are the more detrimental to gene

expression loss of ATRX function is due to the higher likelihood of abnormal structure formation. The mechanisms to this detrimental effect on gene expression could be via (1) physical blocking of RNA polII (Tornaletti, Park-Snyder et al. 2008) (Haeusler, Donnelly et al. 2014), (2) loss of epigenetic memory (Sarkies, Reams et al. 2010, Sarkies, Murat et al. 2011), or (3) gene silencing as a consequence of double-strand breaks induced by aberrant structures (Shanbhag, Rafalska-Metcalf et al. 2010).

In conclusion, this study provides evidence that R-loops might play a role in triggering ATRX targeting to GC-rich tandem repeat, which contributes to our understanding of the mechanism of chromatin remodeler localisation. Future experiments should address which feature that is induced by R-loops directly interact with ATRX. Whilst much work is needed to be done, this work has taken a step further in fully elucidating the mechanism by which ATRX regulates gene expression.

REFERENCES

Aguilera, A. and T. Garcia-Muse (2012). "R loops: from transcription byproducts to threats to genome stability." Mol Cell **46**(2): 115-124.

Ahel, D., Z. Horejsi, N. Wiechens, S. E. Polo, E. Garcia-Wilson, I. Ahel, H. Flynn, M. Skehel, S. C. West, S. P. Jackson, T. Owen-Hughes and S. J. Boulton (2009). "Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1." Science **325**(5945): 1240-1243.

Ahmad, K. and S. Henikoff (2002). "The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly." Mol Cell **9**(6): 1191-1200.

Alexeev, A., A. Mazin and S. C. Kowalczykowski (2003). "Rad54 protein possesses chromatin-remodeling activity stimulated by the Rad51-ssDNA nucleoprotein filament." Nat Struct Biol **10**(3): 182-186.

Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love and E. N. Moudrianakis (1991). "The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix." Proc Natl Acad Sci U S A **88**(22): 10148-10152.

Argentaro, A., J. C. Yang, L. Chapman, M. S. Kowalczyk, R. J. Gibbons, D. R. Higgs, D. Neuhaus and D. Rhodes (2007). "Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX." Proc Natl Acad Sci U S A **104**(29): 11939-11944.

Azvolinsky, A., P. G. Giresi, J. D. Lieb and V. A. Zakian (2009). "Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae." Mol Cell **34**(6): 722-734.

Azzalin, C. M., P. Reichenbach, L. Khoriauli, E. Giulotto and J. Lingner (2007). "Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends." Science **318**(5851): 798-801.

Bagheri-Fam, S., A. Argentaro, T. Svingen, A. N. Combes, A. H. Sinclair, P. Koopman and V. R. Harley (2011). "Defective survival of proliferating Sertoli cells and androgen receptor function in a mouse model of the ATR-X syndrome." Hum Mol Genet **20**(11): 2213-2224.

Balasubramanian, S., L. H. Hurley and S. Neidle (2011). "Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?" Nat Rev Drug Discov **10**(4): 261-275.

Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-395.

Bartholomew, B. (2014). "Regulating the chromatin landscape: structural and mechanistic perspectives." Annu Rev Biochem **83**: 671-696.

Bassett, A. R., S. E. Cooper, A. Ragab and A. A. Travers (2008). "The chromatin remodelling factor dATRX is involved in heterochromatin formation." PLoS One **3**(5): e2099.

Berube, N. G., J. Healy, C. F. Medina, S. Wu, T. Hodgson, M. Jagla and D. J. Picketts (2008). "Patient mutations alter ATRX targeting to PML nuclear bodies." Eur J Hum Genet **16**(2): 192-201.

Berube, N. G., M. Mangelsdorf, M. Jagla, J. Vanderluit, D. Garrick, R. J. Gibbons, D. R. Higgs, R. S. Slack and D. J. Picketts (2005). "The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis." Journal of Clinical Investigation **115**(2): 258-267.

Besnard, E., A. Babled, L. Lapasset, O. Milhavet, H. Parrinello, C. Dantec, J. M. Marin and J. M. Lemaitre (2012). "Unraveling cell type-specific and reprogrammable

human replication origin signatures associated with G-quadruplex consensus motifs." Nat Struct Mol Biol **19**(8): 837-844.
Bhatia, V., S. I. Barroso, M. L. Garcia-Rubio, E. Tumini, E. Herrera-Moyano and A. Aguilera (2014). "BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2." Nature.
Bienz, M. (2006). "The PHD finger, a nuclear protein-interaction domain." Trends Biochem Sci **31**(1): 35-40.
Biffi, G., M. Di Antonio, D. Tannahill and S. Balasubramanian (2014). "Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells." Nat Chem **6**(1): 75-80.
Biffi, G., D. Tannahill, J. McCafferty and S. Balasubramanian (2013). "Quantitative visualization of DNA G-quadruplex structures in human cells." Nat Chem **5**(3): 182-186.
Biffi, G., D. Tannahill, J. Miller, W. J. Howat and S. Balasubramanian (2014). "Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues." PLoS One **9**(7): e102711.
Blackburn, E. H., C. W. Greider and J. W. Szostak (2006). "Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging." Nat Med **12**(10): 1133-1138.
Blosser, T. R., J. G. Yang, M. D. Stone, G. J. Narlikar and X. Zhuang (2009). "Dynamics of nucleosome remodelling by individual ACF complexes." Nature **462**(7276): 1022-1027.
Bochman, M. L., K. Paeschke and V. A. Zakian (2012). "DNA secondary structures: stability and function of G-quadruplex structures." Nat Rev Genet **13**(11): 770-780.
Boguslawski, S. J., D. E. Smith, M. A. Michalak, K. E. Mickelson, C. O. Yehle, W. L. Patterson and R. J. Carrico (1986). "Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids." J Immunol Methods **89**(1): 123-130.
Bohm, F., F. Kappes, I. Scholten, N. Richter, H. Matsuo, R. Knippers and T. Waldmann (2005). "The SAF-box domain of chromatin protein DEK." Nucleic Acids Res **33**(3): 1101-1110.
Bohm, V., A. R. Hieb, A. J. Andrews, A. Gansen, A. Rocker, K. Toth, K. Luger and J. Langowski (2011). "Nucleosome accessibility governed by the dimer/tetramer interface." Nucleic Acids Res **39**(8): 3093-3102.
Boule, J. B. and V. A. Zakian (2007). "The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates." Nucleic Acids Res **35**(17): 5809-5818.
Boyer, L. A., R. R. Latek and C. L. Peterson (2004). "The SANT domain: a unique histone-tail-binding module?" Nat Rev Mol Cell Biol **5**(2): 158-163.
Brewer, B. J., D. Lockshon and W. L. Fangman (1992). "The arrest of replication forks in the rDNA of yeast occurs independently of transcription." Cell **71**(2): 267-276.
Brooks, T. A. and L. H. Hurley (2010). "Targeting MYC Expression through G-Quadruplexes." Genes & cancer **1**(6): 641-649.
Broxson, C., J. Beckett and S. Tornaletti (2011). "Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene." Biochemistry **50**(19): 4162-4172.
Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter (2002). Molecular Biology of the Cell. New York, Garland Science.
Bryant, H. E., E. Petermann, N. Schultz, A. S. Jemth, O. Loseva, N. Issaeva, F. Johansson, S. Fernandez, P. McGlynn and T. Helleday (2009). "PARP is activated at

stalled forks to mediate Mre11-dependent replication restart and recombination." EMBO J **28**(17): 2601-2615.
Buermeyer, A. B., S. M. Deschenes, S. M. Baker and R. M. Liskay (1999). "Mammalian DNA mismatch repair." Annu Rev Genet **33**: 533-564.
Cairns, B. R., Y. J. Kim, M. H. Sayre, B. C. Laurent and R. D. Kornberg (1994). "A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast." Proc Natl Acad Sci U S A **91**(5): 1950-1954.
Cairns, B. R., Y. Lorch, Y. Li, M. Zhang, L. Lacomis, H. Erdjument-Bromage, P. Tempst, J. Du, B. Laurent and R. D. Kornberg (1996). "RSC, an essential, abundant chromatin-remodeling complex." Cell **87**(7): 1249-1260.
Castellano-Pozo, M., J. M. Santos-Pereira, A. G. Rondon, S. Barroso, E. Andujar, M. Perez-Alegre, T. Garcia-Muse and A. Aguilera (2013). "R loops are linked to histone H3 S10 phosphorylation and chromatin condensation." Mol Cell **52**(4): 583-590.
Castillo Bosch, P., S. Segura-Bayona, W. Koole, J. T. van Heteren, J. M. Dewar, M. Tijsterman and P. Knipscheer (2014). "FANCJ promotes DNA synthesis through G-quadruplex structures." EMBO J.
Cerritelli, S. M. and R. J. Crouch (2009). "Ribonuclease H: the enzymes in eukaryotes." FEBS J **276**(6): 1494-1505.
Cesare, A. J. and R. R. Reddel (2010). "Alternative lengthening of telomeres: models, mechanisms and implications." Nat Rev Genet **11**(5): 319-330.
Chakraborty, P. and F. Grosse (2011). "Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes." DNA Repair (Amst) **10**(6): 654-665.
Chan, Y. A., M. J. Aristizabal, P. Y. Lu, Z. Luo, A. Hamza, M. S. Kobor, P. C. Stirling and P. Hieter (2014). "Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip." PLoS Genet **10**(4): e1004288.
Chang, F. T., J. D. McGhie, F. L. Chan, M. C. Tang, M. A. Anderson, J. R. Mann, K. H. Andy Choo and L. H. Wong (2013). "PML bodies provide an important platform for the maintenance of telomeric chromatin integrity in embryonic stem cells." Nucleic Acids Res **41**(8): 4447-4458.
Chatterjee, N., D. Sinha, M. Lemma-Dechassa, S. Tan, M. A. Shogren-Knaak and B. Bartholomew (2011). "Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms." Nucleic Acids Res **39**(19): 8378-8391.
Chaudhuri, J. and F. W. Alt (2004). "Class-switch recombination: interplay of transcription, DNA deamination and DNA repair." Nat Rev Immunol **4**(7): 541-552.
Chu, W. K. and I. D. Hickson (2009). "RecQ helicases: multifunctional genome caretakers." Nat Rev Cancer **9**(9): 644-654.
Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annu Rev Biochem **78**: 273-304.
Clapier, C. R., G. Langst, D. F. Corona, P. B. Becker and K. P. Nightingale (2001). "Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI." Mol Cell Biol **21**(3): 875-883.
Clapier, C. R., K. P. Nightingale and P. B. Becker (2002). "A critical epitope for substrate recognition by the nucleosome remodeling ATPase ISWI." Nucleic Acids Res **30**(3): 649-655.
Clynes, D., C. Jelinska, B. Xella, H. Ayyub, S. Taylor, M. Mitson, C. Z. Bachrati, D. R. Higgs and R. J. Gibbons (2014). "ATRX dysfunction induces replication defects in primary mouse cells." PLoS One **9**(3): e92915.

Cogoi, S., M. Paramasivam, B. Spolaore and L. E. Xodo (2008). "Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins." Nucleic Acids Research **36**(11): 3765-3780.
Cogoi, S. and L. E. Xodo (2006). "G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription." Nucleic acids research **34**(9): 2536-2549.
Colak, D., N. Zaninovic, M. S. Cohen, Z. Rosenwaks, W. Y. Yang, J. Gerhardt, M. D. Disney and S. R. Jaffrey (2014). "Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome." Science **343**(6174): 1002-1005.
Collins, N., R. A. Poot, I. Kukimoto, C. Garcia-Jimenez, G. Dellaire and P. D. Varga-Weisz (2002). "An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin." Nat Genet **32**(4): 627-632.
Collins, R. and X. Cheng (2010). "A case study in cross-talk: the histone lysine methyltransferases G9a and GLP." Nucleic Acids Res **38**(11): 3503-3511.
Connelly, S. and J. L. Manley (1988). "A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II." Genes Dev **2**(4): 440-452.
Conticello, S. G. (2008). "The AID/APOBEC family of nucleic acid mutators." Genome Biol **9**(6): 229.
Cook, P. R. (1999). "The organization of replication and transcription." Science **284**(5421): 1790-1795.
Costanzi, C. and J. R. Pehrson (1998). "Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals." Nature **393**(6685): 599-601.
Costanzi, C. and J. R. Pehrson (2001). "MACROH2A2, a new member of the MACROH2A core histone family." Journal of Biological Chemistry **276**(24): 21776-21784.
Dai, J., T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones and D. Yang (2006). "An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution." Journal of the American Chemical Society **128**(4): 1096-1098.
Daniels, G. A. and M. R. Lieber (1995). "RNA:DNA complex formation upon transcription of immunoglobulin switch regions: Implications for the mechanism and regulation of class switch recombination." Nucleic Acids Research **23**(24): 5006-5011.
De Koning, L., A. Corpet, J. E. Haber and G. Almouzni (2007). "Histone chaperones: an escort network regulating histone traffic." Nat Struct Mol Biol **14**(11): 997-1007.
De La Fuente, R., M. M. Viveiros, K. Wigglesworth and J. J. Eppig (2004). "ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes." Dev Biol **272**(1): 1-14.
Delbarre, E., K. Ivanauskiene, T. Kuntziger and P. Collas (2013). "DAXX-dependent supply of soluble (H3.3-H4) dimers to PML bodies pending deposition into chromatin." Genome Res **23**(3): 440-451.
Dhayalan, A., R. Tamas, I. Bock, A. Tattermusch, E. Dimitrova, S. Kudithipudi, S. Ragozin and A. Jeltsch (2011). "The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9." Hum Mol Genet **20**(11): 2195-2203.

Dominguez-Sanchez, M. S., S. Barroso, B. Gomez-Gonzalez, R. Luna and A. Aguilera (2011). "Genome instability and transcription elongation impairment in human cells depleted of THO/TREX." PLoS Genet **7**(12): e1002386.
Drane, P., K. Ouararhni, A. Depaux, M. Shuaib and A. Hamiche (2010). "The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3." Genes & Development **24**(12): 1253-1265.
Drolet, M., P. Phoenix, R. Menzel, E. Masse, L. F. Liu and R. J. Crouch (1995). "Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I." Proc Natl Acad Sci U S A **92**(8): 3526-3530.
Dunnick, W., G. Z. Hertz, L. Scappino and C. Gritzmacher (1993). "DNA sequences at immunoglobulin switch region recombination sites." Nucleic Acids Res **21**(3): 365-372.
Duquette, M. L., P. Handa, J. A. Vincent, A. F. Taylor and N. Maizels (2004). "Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA." Genes & development **18**(13): 1618-1629.
Duquette, M. L., P. Pham, M. F. Goodman and N. Maizels (2005). "AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation." Oncogene **24**(38): 5791-5798.
Durr, H., C. Korner, M. Muller, V. Hickmann and K. P. Hopfner (2005). "X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA." Cell **121**(3): 363-373.
Eddy, J. and N. Maizels (2006). "Gene function correlates with potential for G4 DNA formation in the human genome." Nucleic Acids Res **34**(14): 3887-3896.
Eddy, J. and N. Maizels (2008). "Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes." Nucleic Acids Res **36**(4): 1321-1333.
Eissenberg, J. C. (2001). "Molecular biology of the chromo domain: an ancient chromatin module comes of age." Gene **275**(1): 19-29.
El Hage, A., S. L. French, A. L. Beyer and D. Tollervey (2010). "Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis." Genes Dev **24**(14): 1546-1558.
Elsasser, S. J., H. Huang, P. W. Lewis, J. W. Chin, C. D. Allis and D. J. Patel (2012). "DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition." Nature **491**(7425): 560-565.
Erdel, F., J. Krug, G. Langst and K. Rippe (2011). "Targeting chromatin remodelers: signals and search mechanisms." Biochim Biophys Acta **1809**(9): 497-508.
Erdel, F. and K. Rippe (2011). "Chromatin remodelling in mammalian cells by ISWI-type complexes - where, when and why?" Febs Journal **278**(19): 3608-3618.
Erdel, F., T. Schubert, C. Marth, G. Langst and K. Rippe (2010). "Human ISWI chromatin-remodeling complexes sample nucleosomes via transient binding reactions and become immobilized at active sites." Proc Natl Acad Sci U S A **107**(46): 19873-19878.
Eustermann, S., J. C. Yang, M. J. Law, R. Amos, L. M. Chapman, C. Jelinska, D. Garrick, D. Clynes, R. J. Gibbons, D. Rhodes, D. R. Higgs and D. Neuhaus (2011). "Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin." Nature structural & molecular biology **18**(7): 777-782.
Farr, C., J. Fantes, P. Goodfellow and H. Cooke (1991). "Functional reintroduction of human telomeres into mammalian cells." Proc Natl Acad Sci U S A **88**(16): 7006-7010.

Fernando, H., S. Sewitz, J. Darot, S. Tavare, J. L. Huppert and S. Balasubramanian (2009). "Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression." Nucleic Acids Res **37**(20): 6716-6722.
Fischle, W., B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki and C. D. Allis (2005). "Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation." Nature **438**(7071): 1116-1122.
Fitzgerald, D. J., C. DeLuca, I. Berger, H. Gaillard, R. Sigrist, K. Schimmele and T. J. Richmond (2004). "Reaction cycle of the yeast Isw2 chromatin remodeling complex." EMBO J **23**(19): 3836-3843.
Flaus, A., D. M. Martin, G. J. Barton and T. Owen-Hughes (2006). "Identification of multiple distinct Snf2 subfamilies with conserved structural motifs." Nucleic Acids Res **34**(10): 2887-2905.
Fry, M. and L. A. Loeb (1999). "Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n." The Journal of biological chemistry **274**(18): 12797-12802.
Fukuda, H., M. Katahira, N. Tsuchiya, Y. Enokizono, T. Sugimura, M. Nagao and H. Nakagama (2002). "Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1." Proc Natl Acad Sci U S A **99**(20): 12685-12690.
Gamble, M. J., K. M. Frizzell, C. Yang, R. Krishnakumar and W. L. Kraus (2010). "The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing." Genes Dev **24**(1): 21-32.
Gan, W., Z. Guan, J. Liu, T. Gui, K. Shen, J. L. Manley and X. Li (2011). "R-loop-mediated genomic instability is caused by impairment of replication fork progression." Genes Dev **25**(19): 2041-2056.
Garrick, D., V. Samara, T. L. McDowell, A. J. Smith, L. Dobbie, D. R. Higgs and R. J. Gibbons (2004). "A conserved truncated isoform of the ATR-X syndrome protein lacking the SWI/SNF-homology domain." Gene **326**: 23-34.
Gellert, M., M. N. Lipsett and D. R. Davies (1962). "Helix formation by guanylic acid." Proc Natl Acad Sci U S A **48**: 2013-2018.
Gerhold, C. B. and S. M. Gasser (2014). "INO80 and SWR complexes: relating structure to function in chromatin remodeling." Trends Cell Biol.
Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea and J. S. Weissman (2003). "Global analysis of protein expression in yeast." Nature **425**(6959): 737-741.
Gibbons, R. (2006). "Alpha thalassaemia-mental retardation, X linked." Orphanet journal of rare diseases **1**: 15.
Gibbons, R. J., T. L. McDowell, S. Raman, D. M. O'Rourke, D. Garrick, H. Ayyub and D. R. Higgs (2000). "Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation." Nature genetics **24**(4): 368-371.
Gibbons, R. J., A. Pellagatti, D. Garrick, W. G. Wood, N. Malik, H. Ayyub, C. Langford, J. Boultwood, J. S. Wainscoat and D. R. Higgs (2003). "Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS)." Nature genetics **34**(4): 446-449.
Gibbons, R. J., T. Wada, C. A. Fisher, N. Malik, M. J. Mitson, D. P. Steensma, A. Fryer, D. R. Goudie, I. D. Krantz and J. Traeger-Synodinos (2008). "Mutations in the chromatin-associated protein ATRX." Hum Mutat **29**(6): 796-802.

Ginno, P. A., Y. W. Lim, P. L. Lott, I. Korf and F. Chedin (2013). "GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination." Genome Res **23**(10): 1590-1600.
Ginno, Paul A., Paul L. Lott, Holly C. Christensen, I. Korf and F. Chédin (2012). "R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters." Molecular Cell **45**(6): 814-825.
Giraldo, R. and D. Rhodes (1994). "The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA." EMBO J **13**(10): 2411-2420.
Goldberg, A. D., L. A. Banaszynski, K. M. Noh, P. W. Lewis, S. J. Elsaesser, S. Stadler, S. Dewell, M. Law, X. Guo, X. Li, D. Wen, A. Chapgier, R. C. DeKelver, J. C. Miller, Y. L. Lee, E. A. Boydston, M. C. Holmes, P. D. Gregory, J. M. Greally, S. Rafii, C. Yang, P. J. Scambler, D. Garrick, R. J. Gibbons, D. R. Higgs, I. M. Cristea, F. D. Urnov, D. Zheng and C. D. Allis (2010). "Distinct factors control histone variant H3.3 localization at specific genomic regions." Cell **140**(5): 678-691.
Goldman, J. A., J. D. Garlick and R. E. Kingston (2010). "Chromatin Remodeling by Imitation Switch (ISWI) Class ATP-dependent Remodelers Is Stimulated by Histone Variant H2A.Z." Journal of Biological Chemistry **285**(7): 4645-4651.
Gomez, D., T. Wenner, B. Brassart, C. Douarre, M. F. O'Donohue, V. El Khoury, K. Shin-Ya, H. Morjani, C. Trentesaux and J. F. Riou (2006). "Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells." J Biol Chem **281**(50): 38721-38729.
Gomez-Gonzalez, B. and A. Aguilera (2007). "Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex." Proceedings of the National Academy of Sciences of the United States of America **104**(20): 8409-8414.
Gonzalez, V., K. Guo, L. Hurley and D. Sun (2009). "Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein." The Journal of biological chemistry **284**(35): 23622-23635.
Gossen, M. and H. Bujard (1992). "Tight control of gene expression in mammalian cells by tetracycline-responsive promoters." Proc Natl Acad Sci U S A **89**(12): 5547-5551.
Grand, C. L., H. Han, R. M. Munoz, S. Weitman, D. D. Von Hoff, L. H. Hurley and D. J. Bearss (2002). "The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo." Mol Cancer Ther **1**(8): 565-573.
Gray, L. T., A. C. Vallur, J. Eddy and N. Maizels (2014). "G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD." Nat Chem Biol **10**(4): 313-318.
Gritzmacher, C. A. (1989). "Molecular aspects of heavy-chain class switching." Crit Rev Immunol **9**(3): 173-200.
Groh, M., M. M. Lufino, R. Wade-Martins and N. Gromak (2014). "R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome." PLoS Genet **10**(5): e1004318.
Grummt, I. and G. Langst (2013). "Epigenetic control of RNA polymerase I transcription in mammalian cells." Biochim Biophys Acta **1829**(3-4): 393-404.
Grune, T., J. Brzeski, A. Eberharter, C. R. Clapier, D. F. Corona, P. B. Becker and C. W. Muller (2003). "Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI." Mol Cell **12**(2): 449-460.
Guan, Y., K. R. Reddy, Q. Zhu, Y. Li, K. Lee, P. Weerasinghe, J. Prchal, G. L. Semenza and N. Jing (2010). "G-rich oligonucleotides inhibit HIF-1alpha and HIF-

2alpha and block tumor growth." Molecular therapy : the journal of the American Society of Gene Therapy **18**(1): 188-197.
Guenatri, M., D. Bailly, C. Maison and G. Almouzni (2004). "Mouse centric and pericentric satellite repeats form distinct functional heterochromatin." J Cell Biol **166**(4): 493-505.
Gupta, R., S. Sharma, J. A. Sommers, Z. Jin, S. B. Cantor and R. M. Brosh, Jr. (2005). "Analysis of the DNA substrate specificity of the human BACH1 helicase associated with breast cancer." J Biol Chem **280**(27): 25450-25460.
Haeusler, A. R., C. J. Donnelly, G. Periz, E. A. Simko, P. G. Shaw, M. S. Kim, N. J. Maragakis, J. C. Troncoso, A. Pandey, R. Sattler, J. D. Rothstein and J. Wang (2014). "C9orf72 nucleotide repeat structures initiate molecular cascades of disease." Nature **507**(7491): 195-200.
Hahn, M. A., X. Wu, A. X. Li, T. Hahn and G. P. Pfeifer (2011). "Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks." PLoS One **6**(4): e18844.
Hamiche, A., J. G. Kang, C. Dennis, H. Xiao and C. Wu (2001). "Histone tails modulate nucleosome mobility and regulate ATP-dependent nucleosome sliding by NURF." Proc Natl Acad Sci U S A **98**(25): 14316-14321.
Hamperl, S. and K. A. Cimprich (2014). "The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability." DNA Repair (Amst) **19**: 84-94.
Haynes, S. R., C. Dollard, F. Winston, S. Beck, J. Trowsdale and I. B. Dawid (1992). "The Bromodomain - a Conserved Sequence Found in Human, Drosophila and Yeast Proteins." Nucleic Acids Research **20**(10): 2603-2603.
Heaphy, C. M., R. F. de Wilde, Y. Jiao, A. P. Klein, B. H. Edil, C. Shi, C. Bettegowda, F. J. Rodriguez, C. G. Eberhart, S. Hebbar, G. J. Offerhaus, R. McLendon, B. A. Rasheed, Y. He, H. Yan, D. D. Bigner, S. M. Oba-Shinjo, S. K. Marie, G. J. Riggins, K. W. Kinzler, B. Vogelstein, R. H. Hruban, A. Maitra, N. Papadopoulos and A. K. Meeker (2011). "Altered telomeres in tumors with ATRX and DAXX mutations." Science **333**(6041): 425.
Henderson, A., Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh, Jr., D. Sen and P. M. Lansdorp (2014). "Detection of G-quadruplex DNA in mammalian cells." Nucleic Acids Res **42**(2): 860-869.
Hickson, I. D., S. L. Davies, J. L. Li, N. C. Levitt, P. Mohaghegh, P. S. North and L. Wu (2001). "Role of the Bloom's syndrome helicase in maintenance of genome stability." Biochemical Society transactions **29**(Pt 2): 201-204.
Hillen, W. and C. Berens (1994). "Mechanisms underlying expression of Tn10 encoded tetracycline resistance." Annu Rev Microbiol **48**: 345-369.
Hofacker, I. L. (2003). "Vienna RNA secondary structure server." Nucleic Acids Res **31**(13): 3429-3431.
Hong, X., G. W. Cadwell and T. Kogoma (1995). "Escherichia coli RecG and RecA proteins in R-loop formation." EMBO J **14**(10): 2385-2392.
Horigome, C., Y. Oma, T. Konishi, R. Schmid, I. Marcomini, M. H. Hauer, V. Dion, M. Harata and S. M. Gasser (2014). "SWR1 and INO80 Chromatin Remodelers Contribute to DNA Double-Strand Break Perinuclear Anchorage Site Choice." Mol Cell.
Huertas, P. and A. Aguilera (2003). "Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination." Mol Cell **12**(3): 711-721.

Huh, M. S., T. Price O'Dea, D. Ouazia, B. C. McKay, G. Parise, R. J. Parks, M. A. Rudnicki and D. J. Picketts (2012). "Compromised genomic integrity impedes muscle growth after Atrx inactivation." J Clin Invest **122**(12): 4412-4423.
Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman and E. K. O'Shea (2003). "Global analysis of protein localization in budding yeast." Nature **425**(6959): 686-691.
Huppert, J. L. (2008). "Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes." Chem Soc Rev **37**(7): 1375-1384.
Huppert, J. L. (2008). "Hunting G-quadruplexes." Biochimie **90**(8): 1140-1148.
Huppert, J. L. (2010). "Structure, location and interactions of G-quadruplexes." The FEBS journal **277**(17): 3452-3458.
Huppert, J. L. and S. Balasubramanian (2005). "Prevalence of quadruplexes in the human genome." Nucleic Acids Res **33**(9): 2908-2916.
Huppert, J. L. and S. Balasubramanian (2007). "G-quadruplexes in promoters throughout the human genome." Nucleic Acids Res **35**(2): 406-413.
Huppert, J. L., A. Bugaut, S. Kumari and S. Balasubramanian (2008). "G-quadruplexes: the beginning and end of UTRs." Nucleic Acids Res **36**(19): 6260-6268.
Indig, F. E., I. Rybanska, P. Karmakar, C. Devulapalli, H. Fu, F. Carrier and V. A. Bohr (2012). "Nucleolin inhibits g4 oligonucleotide unwinding by werner helicase." PloS one **7**(6): e35229-e35229.
Itoh, T. and J. Tomizawa (1980). "Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H." Proc Natl Acad Sci U S A **77**(5): 2450-2454.
Itzhaki, J. E., M. A. Barnett, A. B. MacCarthy, V. J. Buckle, W. R. Brown and A. C. Porter (1992). "Targeted breakage of a human chromosome mediated by cloned human telomeric DNA." Nat Genet **2**(4): 283-287.
Ivanauskiene, K., E. Delbarre, J. McGhie, T. Kuntziger, L. H. Wong and P. Collas (2014). "The PML-associated protein DEK regulates the balance of H3.3 loading on chromatin and is important for telomere integrity." Genome Res.
Ivancic-Bace, I., J. A. Howard and E. L. Bolt (2012). "Tuning in to interference: R-loops and cascade complexes in CRISPR immunity." J Mol Biol **422**(5): 607-616.
Iwase, S., B. Xiang, S. Ghosh, T. Ren, P. W. Lewis, J. C. Cochrane, C. D. Allis, D. J. Picketts, D. J. Patel, H. Li and Y. Shi (2011). "ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome." Nat Struct Mol Biol **18**(7): 769-776.
Jaskelioff, M., S. Van Komen, J. E. Krebs, P. Sung and C. L. Peterson (2003). "Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin." J Biol Chem **278**(11): 9212-9218.
Jenuwein, T. (2006). "The epigenetic magic of histone lysine methylation." FEBS J **273**(14): 3121-3135.
Jeong, K. W., Y. H. Lee and M. R. Stallcup (2009). "Recruitment of the SWI/SNF chromatin remodeling complex to steroid hormone-regulated promoters by nuclear receptor coactivator flightless-I." J Biol Chem **284**(43): 29298-29309.
Jones, D. O., I. G. Cowell and P. B. Singh (2000). "Mammalian chromodomain proteins: their role in genome organisation and expression." Bioessays **22**(2): 124-137.
Jones, M. and A. Rose (2012). "A DOG's View of Fanconi Anemia: Insights from C. elegans." Anemia **2012**: 323721.

Khorasanizadeh, S. (2004). "The nucleosome: from genomic organization to genomic regulation." Cell **116**(2): 259-272.
Kikin, O., L. D'Antonio and P. S. Bagga (2006). "QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences." Nucleic Acids Res **34**(Web Server issue): W676-682.
Kilburn, A. E., M. J. Shea, R. G. Sargent and J. H. Wilson (2001). "Insertion of a telomere repeat sequence into a mammalian gene causes chromosome instability." Molecular and cellular biology **21**(1): 126-135.
Killela, P. J., Z. J. Reitman, Y. Jiao, C. Bettegowda, N. Agrawal, L. A. Diaz, Jr., A. H. Friedman, H. Friedman, G. L. Gallia, B. C. Giovanella, A. P. Grollman, T. C. He, Y. He, R. H. Hruban, G. I. Jallo, N. Mandahl, A. K. Meeker, F. Mertens, G. J. Netto, B. A. Rasheed, G. J. Riggins, T. A. Rosenquist, M. Schiffman, M. Shih Ie, D. Theodorescu, M. S. Torbenson, V. E. Velculescu, T. L. Wang, N. Wentzensen, L. D. Wood, M. Zhang, R. E. McLendon, D. D. Bigner, K. W. Kinzler, B. Vogelstein, N. Papadopoulos and H. Yan (2013). "TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal." Proc Natl Acad Sci U S A **110**(15): 6021-6026.
Kim, J. H., D. S. Choi and H. Kende (2003). "The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis." Plant Journal **36**(1): 94-104.
Kourmouli, N., Y. M. Sun, S. van der Sar, P. B. Singh and J. P. Brown (2005). "Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1." Biochem Biophys Res Commun **337**(3): 901-907.
Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.
Kreuzer, K. N. and J. R. Brister (2010). "Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives." Virol J **7**: 358.
Kruisselbrink, E., V. Guryev, K. Brouwer, D. B. Pontier, E. Cuppen and M. Tijsterman (2008). "Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans." Curr Biol **18**(12): 900-905.
Kubicek, S., R. J. O'Sullivan, E. M. August, E. R. Hickey, Q. Zhang, M. L. Teodoro, S. Rea, K. Mechtler, J. A. Kowalski, C. A. Homon, T. A. Kelly and T. Jenuwein (2007). "Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase." Mol Cell **25**(3): 473-481.
Lallemand-Breitenbach, V. and H. de The (2010). "PML nuclear bodies." Cold Spring Harb Perspect Biol **2**(5): a000661.
Lam, E. Y., D. Beraldi, D. Tannahill and S. Balasubramanian (2013). "G-quadruplex structures are stable and detectable in human genomic DNA." Nat Commun **4**: 1796.
Lan, L., A. Ui, S. Nakajima, K. Hatakeyama, M. Hoshi, R. Watanabe, S. M. Janicki, H. Ogiwara, T. Kohno, S. Kanno and A. Yasui (2010). "The ACF1 complex is required for DNA double-strand break repair in human cells." Mol Cell **40**(6): 976-987.
Langst, L. M. a. G. (2013). Chromatin remodelers and their way of action, Chromatin remodelling, InTech.
Larson, E. D., M. L. Duquette, W. J. Cummings, R. J. Streiff and N. Maizels (2005). "MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions." Curr Biol **15**(5): 470-474.
Law, M. J., K. M. Lower, H. P. J. Voon, J. R. Hughes, D. Garrick, V. Viprakasit, M. Mitson, M. De Gobbi, M. Marra, A. Morris, A. Abbott, S. P. Wilder, S. Taylor, G. M.

Santos, J. Cross, H. Ayyub, S. Jones, J. Ragoussis, D. Rhodes, I. Dunham, D. R. Higgs and R. J. Gibbons (2010). "ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner." Cell **143**(3): 367-378.
Le Douarin, B., A. L. Nielsen, J. M. Garnier, H. Ichinose, F. Jeanmougin, R. Losson and P. Chambon (1996). "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors." EMBO J **15**(23): 6701-6715.
Lechner, M. S., D. C. Schultz, D. Negorev, G. G. Maul and F. J. Rauscher (2005). "The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain." Biochemical and Biophysical Research Communications **331**(4): 929-937.
Lee, H. S., J. H. Park, S. J. Kim, S. J. Kwon and J. Kwon (2010). "A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair." EMBO J **29**(8): 1434-1445.
Leung, J. W., G. Ghosal, W. Wang, X. Shen, J. Wang, L. Li and J. Chen (2013). "Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress." J Biol Chem **288**(9): 6342-6350.
Lewis, P. W., S. J. Elsaesser, K. M. Noh, S. C. Stadler and C. D. Allis (2010). "Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres." Proc Natl Acad Sci U S A **107**(32): 14075-14080.
Li, B., S. G. Pattenden, D. Lee, J. Gutierrez, J. Chen, C. Seidel, J. Gerton and J. L. Workman (2005). "Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling." Proc Natl Acad Sci U S A **102**(51): 18385-18390.
Li, G. and D. Reinberg (2011). "Chromatin higher-order structures and gene regulation." Curr Opin Genet Dev **21**(2): 175-186.
Li, X. and J. L. Manley (2005). "Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability." Cell **122**(3): 365-378.
Lia, G., E. Praly, H. Ferreira, C. Stockdale, Y. C. Tse-Dinh, D. Dunlap, V. Croquette, D. Bensimon and T. Owen-Hughes (2006). "Direct observation of DNA distortion by the RSC complex." Molecular Cell **21**(3): 417-425.
Lipps, H. J. and D. Rhodes (2009). "G-quadruplex structures: in vivo evidence and function." Trends in cell biology **19**(8): 414-422.
London, T. B., L. J. Barber, G. Mosedale, G. P. Kelly, S. Balasubramanian, I. D. Hickson, S. J. Boulton and K. Hiom (2008). "FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts." J Biol Chem **283**(52): 36132-36139.
Loomis, E. W., L. A. Sanz, F. Chedin and P. J. Hagerman (2014). "Transcription-associated R-loop formation across the human FMR1 CGG-repeat region." PLoS Genet **10**(4): e1004294.
Lopes, J., A. Piazza, R. Bermejo, B. Kriegsman, A. Colosio, M. P. Teulade-Fichou, M. Foiani and A. Nicolas (2011). "G-quadruplex-induced instability during leading-strand replication." The EMBO journal **30**(19): 4033-4046.
Lopes, M., M. Foiani and J. M. Sogo (2006). "Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions." Mol Cell **21**(1): 15-27.
Lovejoy, C. A., W. Li, S. Reisenweber, S. Thongthip, J. Bruno, T. de Lange, S. De, J. H. J. Petrini, P. A. Sung, M. Jasin, J. Rosenbluh, Y. Zwang, B. A. Weir, C. Hatton, E.

Ivanova, L. Macconaill, M. Hanna, W. C. Hahn, N. F. Lue, R. R. Reddel, Y. Jiao, K. Kinzler, B. Vogelstein, N. Papadopoulos and A. K. Meeker (2012). "Loss of ATRX, Genome Instability, and an Altered DNA Damage Response Are Hallmarks of the Alternative Lengthening of Telomeres Pathway." PLoS genetics **8**(7): e1002772-e1002772.
Luger, K., M. L. Dechassa and D. J. Tremethick (2012). "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" Nat Rev Mol Cell Biol **13**(7): 436-447.
Luke, B., A. Panza, S. Redon, N. Iglesias, Z. Li and J. Lingner (2008). "The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae." Molecular cell **32**(4): 465-477.
Luo, R. X. and D. C. Dean (1999). "Chromatin remodeling and transcriptional regulation." J Natl Cancer Inst **91**(15): 1288-1294.
Lusser, A., D. L. Urwin and J. T. Kadonaga (2005). "Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly." Nat Struct Mol Biol **12**(2): 160-166.
Maison, C., D. Bailly, D. Roche, R. Montes de Oca, A. V. Probst, I. Vassias, F. Dingli, B. Lombard, D. Loew, J. P. Quivy and G. Almouzni (2011). "SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin." Nat Genet **43**(3): 220-227.
Manelyte, L., R. Strohner, T. Gross and G. Langst (2014). "Chromatin targeting signals, nucleosome positioning mechanism and non-coding RNA-mediated regulation of the chromatin remodeling complex NoRC." PLoS Genet **10**(3): e1004157.
Marfella, C. G. and A. N. Imbalzano (2007). "The Chd family of chromatin remodelers." Mutat Res **618**(1-2): 30-40.
Margueron, R. and D. Reinberg (2010). "Chromatin structure and the inheritance of epigenetic information." Nat Rev Genet **11**(4): 285-296.
Maze, I., K. M. Noh, A. A. Soshnev and C. D. Allis (2014). "Every amino acid matters: essential contributions of histone variants to mammalian development and disease." Nat Rev Genet **15**(4): 259-271.
McDowell, T. L., R. J. Gibbons, H. Sutherland, D. M. O'Rourke, W. A. Bickmore, A. Pombo, H. Turley, K. Gatter, D. J. Picketts, V. J. Buckle, L. Chapman, D. Rhodes and D. R. Higgs (1999). "Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes." Proceedings of the National Academy of Sciences of the United States of America **96**(24): 13983-13988.
Mergny, J. L., A. T. Phan and L. Lacroix (1998). "Following G-quartet formation by UV-spectroscopy." FEBS Lett **435**(1): 74-78.
Mermoud, J. E., C. Costanzi, J. R. Pehrson and N. Brockdorff (1999). "Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation." J Cell Biol **147**(7): 1399-1408.
Mischo, H. E., B. Gomez-Gonzalez, P. Grzechnik, A. G. Rondon, W. Wei, L. Steinmetz, A. Aguilera and N. J. Proudfoot (2011). "Yeast Sen1 helicase protects the genome from transcription-associated instability." Mol Cell **41**(1): 21-32.
Mohaghegh, P. and I. D. Hickson (2001). "DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders." Human molecular genetics **10**(7): 741-746.
Mohaghegh, P., J. K. Karow, R. M. Brosh, Jr., V. A. Bohr and I. D. Hickson (2001). "The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases." Nucleic acids research **29**(13): 2843-2849.

Mokry, M., P. Hatzis, J. Schuijers, N. Lansu, F. P. Ruzius, H. Clevers and E. Cuppen (2012). "Integrated genome-wide analysis of transcription factor occupancy, RNA polymerase II binding and steady-state RNA levels identify differentially regulated functional gene classes." Nucleic Acids Res **40**(1): 148-158.
Morettini, S., M. Tribus, A. Zeilner, J. Sebald, B. Campo-Fernandez, G. Scheran, H. Worle, V. Podhraski, D. V. Fyodorov and A. Lusser (2011). "The chromodomains of CHD1 are critical for enzymatic activity but less important for chromatin localization." Nucleic Acids Res **39**(8): 3103-3115.
Morris, M. J., K. L. Wingate, J. Silwal, T. C. Leeper and S. Basu (2012). "The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells." Nucleic Acids Res **40**(9): 4137-4145.
Muller, S., D. A. Sanders, M. Di Antonio, S. Matsis, J. F. Riou, R. Rodriguez and S. Balasubramanian (2012). "Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells." Org Biomol Chem **10**(32): 6537-6546.
Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai and T. Honjo (2000). "Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme." Cell **102**(5): 553-563.
Murawska, M. and A. Brehm (2011). "CHD chromatin remodelers and the transcription cycle." Transcription **2**(6): 244-253.
Narlikar, G. J., R. Sundaramoorthy and T. Owen-Hughes (2013). "Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes." Cell **154**(3): 490-503.
Naughton, C., N. Avlonitis, S. Corless, J. G. Prendergast, I. K. Mati, P. P. Eijk, S. L. Cockroft, M. Bradley, B. Ylstra and N. Gilbert (2013). "Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures." Nat Struct Mol Biol **20**(3): 387-395.
Neigeborn, L. and M. Carlson (1984). "Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae." Genetics **108**(4): 845-858.
Nergadze, S. G., B. O. Farnung, H. Wischnewski, L. Khoriauli, V. Vitelli, R. Chawla, E. Giulotto and C. M. Azzalin (2009). "CpG-island promoters drive transcription of human telomeres." RNA **15**(12): 2186-2194.
Nguyen, V. Q., A. Ranjan, F. Stengel, D. Wei, R. Aebersold, C. Wu and A. E. Leschziner (2013). "Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1." Cell **154**(6): 1220-1231.
Nielsen, P. R., D. Nietlispach, H. R. Mott, J. Callaghan, A. Bannister, T. Kouzarides, A. G. Murzin, N. V. Murzina and E. D. Laue (2002). "Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9." Nature **416**(6876): 103-107.
O'Sullivan, R. J. and J. Karlseder (2010). "Telomeres: protecting chromosomes against genome instability." Nat Rev Mol Cell Biol **11**(3): 171-181.
Okada, M., K. Okawa, T. Isobe and T. Fukagawa (2009). "CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1." Mol Biol Cell **20**(18): 3986-3995.
Oki, M., H. Aihara and T. Ito (2007). "Role of histone phosphorylation in chromatin dynamics and its implications in diseases." Subcell Biochem **41**: 319-336.
Paeschke, K., J. A. Capra and V. A. Zakian (2011). "DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase." Cell **145**(5): 678-691.

Paeschke, K., S. Juranek, T. Simonsson, A. Hempel, D. Rhodes and H. J. Lipps (2008). "Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates." Nat Struct Mol Biol **15**(6): 598-604.
Papamichos-Chronakis, M., S. Watanabe, O. J. Rando and C. L. Peterson (2011). "Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity." Cell **144**(2): 200-213.
Paramasivam, M., A. Membrino, S. Cogoi, H. Fukuda, H. Nakagama and L. E. Xodo (2009). "Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription." Nucleic Acids Res **37**(9): 2841-2853.
Parkinson, G. N., M. P. Lee and S. Neidle (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA." Nature **417**(6891): 876-880.
Paro, R. and D. S. Hogness (1991). "The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila." Proc Natl Acad Sci U S A **88**(1): 263-267.
Partensky, P. D. and G. J. Narlikar (2009). "Chromatin remodelers act globally, sequence positions nucleosomes locally." J Mol Biol **391**(1): 12-25.
Pehrson, J. R. and V. A. Fried (1992). "MacroH2A, a core histone containing a large nonhistone region." Science **257**(5075): 1398-1400.
Perpelescu, M., N. Nozaki, C. Obuse, H. Yang and K. Yoda (2009). "Active establishment of centromeric CENP-A chromatin by RSF complex." J Cell Biol **185**(3): 397-407.
Petty, E. and L. Pillus (2013). "Balancing chromatin remodeling and histone modifications in transcription." Trends Genet **29**(11): 621-629.
Piazza, A., J. B. Boule, J. Lopes, K. Mingo, E. Largy, M. P. Teulade-Fichou and A. Nicolas (2010). "Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae." Nucleic Acids Res **38**(13): 4337-4348.
Picketts, D. J., A. O. Tastan, D. R. Higgs and R. J. Gibbons (1998). "Comparison of the human and murine ATRX gene identifies highly conserved, functionally important domains." Mamm Genome **9**(5): 400-403.
Polo, S. E., A. Kaidi, L. Baskcomb, Y. Galanty and S. P. Jackson (2010). "Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4." Embo Journal **29**(18): 3130-3139.
Powell, W. T., R. L. Coulson, M. L. Gonzales, F. K. Crary, S. S. Wong, S. Adams, R. A. Ach, P. Tsang, N. A. Yamada, D. H. Yasui, F. Chedin and J. M. LaSalle (2013). "R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation." Proc Natl Acad Sci U S A **110**(34): 13938-13943.
Probst, A. V. and G. Almouzni (2008). "Pericentric heterochromatin: dynamic organization during early development in mammals." Differentiation **76**(1): 15-23.
Proudfoot, N. J., A. Gil and T. Maniatis (1982). "The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene." Cell **31**(3 Pt 2): 553-563.
Qin, Y. and L. H. Hurley (2008). "
." Biochimie **90**(8): 1149-1171.
Ratnakumar, K., L. F. Duarte, G. LeRoy, D. Hasson, D. Smeets, C. Vardabasso, C. Bonisch, T. Zeng, B. Xiang, D. Y. Zhang, H. Li, X. Wang, S. B. Hake, L. Schermelleh, B. A. Garcia and E. Bernstein (2012). "ATRX-mediated chromatin association of histone variant macroH2A1 regulates -globin expression." Genes & Development **26**(5): 433-438.

Reaban, M. E. and J. A. Griffin (1990). "Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region." Nature **348**(6299): 342-344.
Reaban, M. E., J. Lebowitz and J. A. Griffin (1994). "Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region." J Biol Chem **269**(34): 21850-21857.
Reddy, B. A., P. K. Bajpe, A. Bassett, Y. M. Moshkin, E. Kozhevnikova, K. Bezstarosti, J. A. Demmers, A. A. Travers and C. P. Verrijzer (2010). "Drosophila transcription factor Tramtrack69 binds MEP1 to recruit the chromatin remodeler NuRD." Mol Cell Biol **30**(21): 5234-5244.
Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A. G. Ugazio, N. Brousse, M. Muramatsu, L. D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer and A. Durandy (2000). "Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)." Cell **102**(5): 565-575.
Ribeyre, C., J. Lopes, J.-B. Boulé, A. Piazza, A. Guédin, V. A. Zakian, J.-L. Mergny and A. Nicolas (2009). "The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo." PLoS genetics **5**(5): e1000475-e1000475.
Rippe, K., A. Schrader, P. Riede, R. Strohner, E. Lehmann and G. Langst (2007). "DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes." Proc Natl Acad Sci U S A **104**(40): 15635-15640.
Ritchie, K., C. Seah, J. Moulin, C. Isaac, F. Dick and N. G. Berube (2008). "Loss of ATRX leads to chromosome cohesion and congression defects." The Journal of cell biology **180**(2): 315-324.
Roberts, R. W. and D. M. Crothers (1992). "Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition." Science **258**(5087): 1463-1466.
Robinson, K. M. and M. C. Schultz (2003). "Replication-independent assembly of nucleosome arrays in a novel yeast chromatin reconstitution system involves antisilencing factor Asf1p and chromodomain protein Chd1p." Mol Cell Biol **23**(22): 7937-7946.
Robison, J. G., J. Elliott, K. Dixon and G. G. Oakley (2004). "Replication protein A and the Mre11.Rad50.Nbs1 complex co-localize and interact at sites of stalled replication forks." J Biol Chem **279**(33): 34802-34810.
Rodriguez, R., K. M. Miller, J. V. Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian and S. P. Jackson (2012). "Small-molecule-induced DNA damage identifies alternative DNA structures in human genes." Nature chemical biology **8**(3): 301-310.
Rodriguez, R., S. Muller, J. A. Yeoman, C. Trentesaux, J. F. Riou and S. Balasubramanian (2008). "A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres." Journal of the American Chemical Society **130**(47): 15758-+.
Roy, D. and M. R. Lieber (2009). "G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter." Mol Cell Biol **29**(11): 3124-3133.
Roy, D., Z. Zhang, Z. Lu, C. L. Hsieh and M. R. Lieber (2010). "Competition between the RNA transcript and the nontemplate DNA strand during R-loop

formation in vitro: a nick can serve as a strong R-loop initiation site." Mol Cell Biol **30**(1): 146-159.
Saha, A., J. Wittmeyer and B. R. Cairns (2002). "Chromatin remodeling by RSC involves ATP-dependent DNA translocation." Genes Dev **16**(16): 2120-2134.
Saha, A., J. Wittmeyer and B. R. Cairns (2006). "Chromatin remodelling: the industrial revolution of DNA around histones." Nat Rev Mol Cell Biol **7**(6): 437-447.
Salinas-Rios, V., B. P. Belotserkovskii and P. C. Hanawalt (2011). "DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability." Nucleic Acids Res **39**(17): 7444-7454.
Salisbury, J., K. W. Hutchison and J. H. Graber (2006). "A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif." BMC Genomics **7**: 55.
Sanders, C. M. (2010). "Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity." Biochem J **430**(1): 119-128.
Sarkies, P., P. Murat, L. G. Phillips, K. J. Patel, S. Balasubramanian and J. E. Sale (2011). "FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA." Nucleic acids research.
Sarkies, P., C. Reams, L. J. Simpson and J. E. Sale (2010). "Epigenetic instability due to defective replication of structured DNA." Molecular cell **40**(5): 703-713.
Sarma, K. and D. Reinberg (2005). "Histone variants meet their match." Nat Rev Mol Cell Biol **6**(2): 139-149.
Schaffitzel, C., I. Berger, J. Postberg, J. Hanes, H. J. Lipps and A. Pluckthun (2001). "In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei." Proceedings of the National Academy of Sciences of the United States of America **98**(15): 8572-8577.
Schiavone, D., G. Guilbaud, P. Murat, C. Papadopoulou, P. Sarkies, M. N. Prioleau, S. Balasubramanian and J. E. Sale (2014). "Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells." EMBO J.
Schneiderman, J. I., G. A. Orsi, K. T. Hughes, B. Loppin and K. Ahmad (2012). "Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant." Proc Natl Acad Sci U S A **109**(48): 19721-19726.
Schneiderman, J. I., A. Sakai, S. Goldstein and K. Ahmad (2009). "The XNP remodeler targets dynamic chromatin in Drosophila." Proc Natl Acad Sci U S A **106**(34): 14472-14477.
Schoeftner, S. and M. A. Blasco (2008). "Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II." Nat Cell Biol **10**(2): 228-236.
Schwab, R. A., J. Nieminuszczy, K. Shin-ya and W. Niedzwiedz (2013). "FANCJ couples replication past natural fork barriers with maintenance of chromatin structure." J Cell Biol **201**(1): 33-48.
Schwartzentruber, J., A. Korshunov, X. Y. Liu, D. T. Jones, E. Pfaff, K. Jacob, D. Sturm, A. M. Fontebasso, D. A. Quang, M. Tonjes, V. Hovestadt, S. Albrecht, M. Kool, A. Nantel, C. Konermann, A. Lindroth, N. Jager, T. Rausch, M. Ryzhova, J. O. Korbel, T. Hielscher, P. Hauser, M. Garami, A. Klekner, L. Bognar, M. Ebinger, M. U. Schuhmann, W. Scheurlen, A. Pekrun, M. C. Fruhwald, W. Roggendorf, C. Kramm, M. Durken, J. Atkinson, P. Lepage, A. Montpetit, M. Zakrzewska, K. Zakrzewski, P. P. Liberski, Z. Dong, P. Siegel, A. E. Kulozik, M. Zapatka, A. Guha, D. Malkin, J. Felsberg, G. Reifenberger, A. von Deimling, K. Ichimura, V. P. Collins, H. Witt, T. Milde, O. Witt, C. Zhang, P. Castelo-Branco, P. Lichter, D. Faury, U. Tabori, C. Plass, J. Majewski, S. M. Pfister and N. Jabado (2012). "Driver mutations

in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma." Nature **482**(7384): 226-231.
Sen, D. and W. Gilbert (1988). "Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis." Nature **334**(6180): 364-366.
Sen, P., S. Ghosh, B. F. Pugh and B. Bartholomew (2011). "A new, highly conserved domain in Swi2/Snf2 is required for SWI/SNF remodeling." Nucleic Acids Research **39**(21): 9155-9166.
Sfeir, A., S. T. Kosiyatrakul, D. Hockemeyer, S. L. MacRae, J. Karlseder, C. L. Schildkraut and T. de Lange (2009). "Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication." Cell **138**(1): 90-103.
Shanbhag, N. M., I. U. Rafalska-Metcalf, C. Balane-Bolivar, S. M. Janicki and R. A. Greenberg (2010). "ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks." Cell **141**(6): 970-981.
Shaw, N. N. and D. P. Arya (2008). "Recognition of the unique structure of DNA:RNA hybrids." Biochimie **90**(7): 1026-1039.
Shay, J. W. and S. Bacchetti (1997). "A survey of telomerase activity in human cancer." Eur J Cancer **33**(5): 787-791.
Shay, J. W., R. R. Reddel and W. E. Wright (2012). "Cancer. Cancer and telomeres--an ALTernative to telomerase." Science **336**(6087): 1388-1390.
Shen, X., G. Mizuguchi, A. Hamiche and C. Wu (2000). "A chromatin remodelling complex involved in transcription and DNA processing." Nature **406**(6795): 541-544.
Skourti-Stathaki, K. and N. J. Proudfoot (2013). "Histone 3 s10 phosphorylation: "caught in the R loop!"." Mol Cell **52**(4): 470-472.
Skourti-Stathaki, K. and N. J. Proudfoot (2014). "A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression." Genes Dev **28**(13): 1384-1396.
Skourti-Stathaki, K., N. J. Proudfoot and N. Gromak (2011). "Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination." Mol Cell **42**(6): 794-805.
Smith, J. S., Q. Chen, L. A. Yatsunyk, J. M. Nicoludis, M. S. Garcia, R. Kranaster, S. Balasubramanian, D. Monchaud, M. P. Teulade-Fichou, L. Abramowitz, D. C. Schultz and F. B. Johnson (2011). "Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae." Nat Struct Mol Biol **18**(4): 478-485.
Solovei, I., E. R. Gaginskaya and H. C. Macgregor (1994). "The arrangement and transcription of telomere DNA sequences at the ends of lampbrush chromosomes of birds." Chromosome Res **2**(6): 460-470.
Steensma, D. P., D. R. Higgs, C. A. Fisher and R. J. Gibbons (2004). "Acquired somatic ATRX mutations in myelodysplastic syndrome associated with alpha thalassemia (ATMDS) convey a more severe hematologic phenotype than germline ATRX mutations." Blood **103**(6): 2019-2026.
Stirling, P. C., Y. A. Chan, S. W. Minaker, M. J. Aristizabal, I. Barrett, P. Sipahimalani, M. S. Kobor and P. Hieter (2012). "R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants." Genes Dev **26**(2): 163-175.
Stockdale, C., A. Flaus, H. Ferreira and T. Owen-Hughes (2006). "Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes." J Biol Chem **281**(24): 16279-16288.
Strohner, R., A. Nemeth, P. Jansa, U. Hofmann-Rohrer, R. Santoro, G. Langst and I. Grummt (2001). "NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines." EMBO J **20**(17): 4892-4900.

Strohner, R., M. Wachsmuth, K. Dachauer, J. Mazurkiewicz, J. Hochstatter, K. Rippe and G. Langst (2005). "A 'loop recapture' mechanism for ACF-dependent nucleosome remodeling." Nat Struct Mol Biol **12**(8): 683-690.
Sudarsanam, P. and F. Winston (2000). "The Swi/Snf family - nucleosome-remodeling complexes and transcriptional control." Trends in Genetics **16**(8): 345-351.
Sun, D. and L. H. Hurley (2009). "The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression." Journal of medicinal chemistry **52**(9): 2863-2874.
Sun, D., W. J. Liu, K. Guo, J. J. Rusche, S. Ebbinghaus, V. Gokhale and L. H. Hurley (2008). "The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents." Molecular cancer therapeutics **7**(4): 880-889.
Sun, H., R. J. Bennett and N. Maizels (1999). "The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs." Nucleic Acids Res **27**(9): 1978-1984.
Sun, H., J. K. Karow, I. D. Hickson and N. Maizels (1998). "The Bloom's syndrome helicase unwinds G4 DNA." The Journal of biological chemistry **273**(42): 27587-27592.
Sun, H., J. Wu, P. Wickramasinghe, S. Pal, R. Gupta, A. Bhattacharyya, F. J. Agosto-Perez, L. C. Showe, T. H. Huang and R. V. Davuluri (2011). "Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq." Nucleic Acids Res **39**(1): 190-201.
Sun, H., A. Yabuki and N. Maizels (2001). "A human nuclease specific for G4 DNA." Proc Natl Acad Sci U S A **98**(22): 12444-12449.
Sun, Q. W., T. Csorba, K. Skourti-Stathaki, N. J. Proudfoot and C. Dean (2013). "R-Loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus." Science **340**(6132): 619-621.
Szerlong, H., K. Hinata, R. Viswanathan, H. Erdjument-Bromage, P. Tempst and B. R. Cairns (2008). "The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases." Nat Struct Mol Biol **15**(5): 469-476.
Tang, L., E. Nogales and C. Ciferri (2010). "Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription." Prog Biophys Mol Biol **102**(2-3): 122-128.
Tang, M. C., S. A. Jacobs, L. H. Wong and J. R. Mann (2013). "Conditional allelic replacement applied to genes encoding the histone variant H3.3 in the mouse." Genesis **51**(2): 142-146.
Tarsounas, M. and M. Tijsterman (2013). "Genomes and G-quadruplexes: for better or for worse." J Mol Biol **425**(23): 4782-4789.
Taverna, S. D., H. Li, A. J. Ruthenburg, C. D. Allis and D. J. Patel (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-1040.
Teves, S. S. and S. Henikoff (2014). "Transcription-generated torsional stress destabilizes nucleosomes." Nat Struct Mol Biol **21**(1): 88-94.
Thiru, A., D. Nietlispach, H. R. Mott, M. Okuwaki, D. Lyon, P. R. Nielsen, M. Hirshberg, A. Verreault, N. V. Murzina and E. D. Laue (2004). "Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin." EMBO J **23**(3): 489-499.

Tian, M. and F. W. Alt (2000). "Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro." J Biol Chem **275**(31): 24163-24172.
Todd, A. K. (2007). "Bioinformatics approaches to quadruplex sequence location." Methods **43**(4): 246-251.
Todd, A. K., M. Johnston and S. Neidle (2005). "Highly prevalent putative quadruplex sequence motifs in human DNA." Nucleic Acids Res **33**(9): 2901-2907.
Tornaletti, S., S. Park-Snyder and P. C. Hanawalt (2008). "G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II." J Biol Chem **283**(19): 12756-12762.
Tosi, A., C. Haas, F. Herzog, A. Gilmozzi, O. Berninghausen, C. Ungewickell, C. B. Gerhold, K. Lakomek, R. Aebersold, R. Beckmann and K. P. Hopfner (2013). "Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex." Cell **154**(6): 1207-1219.
Tran, H. G., D. J. Steger, V. R. Iyer and A. D. Johnson (2000). "The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor." EMBO J **19**(10): 2323-2331.
Tsukuda, T., A. B. Fleming, J. A. Nickoloff and M. A. Osley (2005). "Chromatin remodelling at a DNA double-strand break site in Saccharomyces cerevisiae." Nature **438**(7066): 379-383.
Tuduri, S., L. Crabbe, C. Conti, H. Tourriere, H. Holtgreve-Grez, A. Jauch, V. Pantesco, J. De Vos, A. Thomas, C. Theillet, Y. Pommier, J. Tazi, A. Coquelle and P. Pasero (2009). "Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription." Nat Cell Biol **11**(11): 1315-1324.
Udugama, M., A. Sabri and B. Bartholomew (2011). "The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor." Mol Cell Biol **31**(4): 662-673.
Uringa, E. J., J. L. Youds, K. Lisaingo, P. M. Lansdorp and S. J. Boulton (2011). "RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination." Nucleic Acids Res **39**(5): 1647-1655.
Valadez-Graham, V., Y. Yoshioka, O. Velazquez, A. Kawamori, M. Vazquez, A. Neumann, M. Yamaguchi and M. Zurita (2012). "XNP/dATRX interacts with DREF in the chromatin to regulate gene expression." Nucleic Acids Res **40**(4): 1460-1474.
Valton, A. L., V. Hassan-Zadeh, I. Lema, N. Boggetto, P. Alberti, C. Saintome, J. F. Riou and M. N. Prioleau (2014). "G4 motifs affect origin positioning and efficiency in two vertebrate replicators." EMBO J **33**(7): 732-746.
van Attikum, H., O. Fritsch and S. M. Gasser (2007). "Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks." EMBO J **26**(18): 4113-4125.
van Vugt, J. J., M. de Jager, M. Murawska, A. Brehm, J. van Noort and C. Logie (2009). "Multiple aspects of ATP-dependent nucleosome translocation by RSC and Mi-2 are directed by the underlying DNA sequence." PLoS One **4**(7): e6345.
Vannier, J. B., V. Pavicic-Kaltenbrunner, M. I. Petalcorin, H. Ding and S. J. Boulton (2012). "RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity." Cell **149**(4): 795-806.
Vannier, J. B., S. Sandhu, M. I. R. Petalcorin, X. L. Wu, Z. Nabi, H. Ding and S. J. Boulton (2013). "RTEL1 Is a Replisome-Associated Helicase That Promotes Telomere and Genome-Wide Replication." Science **342**(6155): 239-242.

Varga-Weisz, P. D. (2010). "Insights into how chromatin remodeling factors find their target in the nucleus." Proceedings of the National Academy of Sciences of the United States of America **107**(46): 19611-19612.
Veloso, A., K. S. Kirkconnell, B. Magnuson, B. Biewen, M. T. Paulsen, T. E. Wilson and M. Ljungman (2014). "Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications." Genome Res **24**(6): 896-905.
Verdone, L., M. Caserta and E. Di Mauro (2005). "Role of histone acetylation in the control of gene expression." Biochem Cell Biol **83**(3): 344-353.
Vignali, M., A. H. Hassan, K. E. Neely and J. L. Workman (2000). "ATP-dependent chromatin-remodeling complexes." Mol Cell Biol **20**(6): 1899-1910.
Villard, L., M. Fontes and J. J. Ewbank (1999). "Characterization of xnp-1, a Caenorhabditis elegans gene similar to the human XNP/ATR-X gene." Gene **236**(1): 13-19.
Villeponteau, B. and H. G. Martinson (1987). "Gamma rays and bleomycin nick DNA and reverse the DNase I sensitivity of beta-globin gene chromatin in vivo." Mol Cell Biol **7**(5): 1917-1924.
Wahba, L., J. D. Amon, D. Koshland and M. Vuica-Ross (2011). "RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability." Mol Cell **44**(6): 978-988.
Wahba, L., S. K. Gore and D. Koshland (2013). "The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability." Elife **2**: e00505.
Waldmann, T., I. Scholten, F. Kappes, H. G. Hu and R. Knippers (2004). "The DEK protein--an abundant and ubiquitous constituent of mammalian chromatin." Gene **343**(1): 1-9.
Wang, J. C. (2002). "Cellular roles of DNA topoisomerases: a molecular perspective." Nat Rev Mol Cell Biol **3**(6): 430-440.
Wang, J. C. and A. S. Lynch (1993). "Transcription and DNA supercoiling." Curr Opin Genet Dev **3**(5): 764-768.
Webba da Silva, M. (2007). "NMR methods for studying quadruplex nucleic acids." Methods **43**(4): 264-277.
Weidemann, T., M. Wachsmuth, T. A. Knoch, G. Muller, W. Waldeck and J. Langowski (2003). "Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging." J Mol Biol **334**(2): 229-240.
Weider, M., M. Kuspert, M. Bischof, M. R. Vogl, J. Hornig, K. Loy, T. Kosian, J. Muller, S. Hillgartner, E. R. Tamm, D. Metzger and M. Wegner (2012). "Chromatin-Remodeling Factor Brg1 Is Required for Schwann Cell Differentiation and Myelination." Developmental Cell **23**(1): 193-201.
West, S., N. J. Proudfoot and M. J. Dye (2008). "Molecular dissection of mammalian RNA polymerase II transcriptional termination." Mol Cell **29**(5): 600-610.
Westover, K. D., D. A. Bushnell and R. D. Kornberg (2004). "Structural basis of transcription: separation of RNA from DNA by RNA polymerase II." Science **303**(5660): 1014-1016.
Whitehouse, I., C. Stockdale, A. Flaus, M. D. Szczelkun and T. Owen-Hughes (2003). "Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme." Mol Cell Biol **23**(6): 1935-1945.

Whitelaw, E. and N. Proudfoot (1986). "Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene." EMBO J **5**(11): 2915-2922.
Williamson, J. R., M. K. Raghuraman and T. R. Cech (1989). "Monovalent cation-induced structure of telomeric DNA: the G-quartet model." Cell **59**(5): 871-880.
Wilson, B. G. and C. W. Roberts (2011). "SWI/SNF nucleosome remodellers and cancer." Nat Rev Cancer **11**(7): 481-492.
Wolfe, A. L., K. Singh, Y. Zhong, P. Drewe, V. K. Rajasekhar, V. R. Sanghvi, K. J. Mavrakis, M. Jiang, J. E. Roderick, J. Van der Meulen, J. H. Schatz, C. M. Rodrigo, C. Zhao, P. Rondou, E. de Stanchina, J. Teruya-Feldstein, M. A. Kelliher, F. Speleman, J. A. Porco, Jr., J. Pelletier, G. Ratsch and H. G. Wendel (2014). "RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer." Nature **513**(7516): 65-70.
Wong, L. H., J. D. McGhie, M. Sim, M. A. Anderson, S. Ahn, R. D. Hannan, A. J. George, K. A. Morgan, J. R. Mann and K. H. Choo (2010). "ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells." Genome Res **20**(3): 351-360.
Woodford, K. J., R. M. Howell and K. Usdin (1994). "A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes." J Biol Chem **269**(43): 27029-27035.
Workman, J. L. and R. E. Kingston (1998). "Alteration of nucleosome structure as a mechanism of transcriptional regulation." Annual Review of Biochemistry **67**: 545-579.
Wu, W. H., S. Alami, E. Luk, C. H. Wu, S. Sen, G. Mizuguchi, D. Wei and C. Wu (2005). "Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange." Nat Struct Mol Biol **12**(12): 1064-1071.
Wu, X. and N. Maizels (2001). "Substrate-specific inhibition of RecQ helicase." Nucleic Acids Res **29**(8): 1765-1771.
Wu, Y. and R. M. Brosh, Jr. (2010). "G-quadruplex nucleic acids and human disease." The FEBS journal **277**(17): 3470-3488.
Wu, Y., K. Shin-ya and R. M. Brosh, Jr. (2008). "FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability." Mol Cell Biol **28**(12): 4116-4128.
Wu, Y., J. A. Sommers, I. Khan, J. P. de Winter and R. M. Brosh, Jr. (2012). "Biochemical characterization of Warsaw breakage syndrome helicase." J Biol Chem **287**(2): 1007-1021.
Wysocka, J., T. Swigut, H. Xiao, T. A. Milne, S. Y. Kwon, J. Landry, M. Kauer, A. J. Tackett, B. T. Chait, P. Badenhorst, C. Wu and C. D. Allis (2006). "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling." Nature **442**(7098): 86-90.
Xiao, A., H. Li, D. Shechter, S. H. Ahn, L. A. Fabrizio, H. Erdjument-Bromage, S. Ishibe-Murakami, B. Wang, P. Tempst, K. Hofmann, D. J. Patel, S. J. Elledge and C. D. Allis (2009). "WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity." Nature **457**(7225): 57-62.
Xu, B. J. and D. A. Clayton (1996). "RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: An implication for RNA-DNA hybrids serving as primers." Embo Journal **15**(12): 3135-3143.
Xue, Y., R. Gibbons, Z. Yan, D. Yang, T. L. McDowell, S. Sechi, J. Qin, S. Zhou, D. Higgs and W. Wang (2003). "The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear

bodies." Proceedings of the National Academy of Sciences of the United States of America **100**(19): 10635-10640.
Yadon, A. N., B. N. Singh, M. Hampsey and T. Tsukiyama (2013). "DNA looping facilitates targeting of a chromatin remodeling enzyme." Mol Cell **50**(1): 93-103.
Yamada, K., T. D. Frouws, B. Angst, D. J. Fitzgerald, C. DeLuca, K. Schimmele, D. F. Sargent and T. J. Richmond (2011). "Structure and mechanism of the chromatin remodelling factor ISW1a." Nature **472**(7344): 448-453.
Yang, J. G., T. S. Madrid, E. Sevastopoulos and G. J. Narlikar (2006). "The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing." Nature Structural & Molecular Biology **13**(12): 1078-1083.
Yu, K., F. Chedin, C. L. Hsieh, T. E. Wilson and M. R. Lieber (2003). "R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells." Nat Immunol **4**(5): 442-451.
Yudkovsky, N., C. Logie, S. Hahn and C. L. Peterson (1999). "Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators." Genes Dev **13**(18): 2369-2374.
Zamiri, B., K. Reddy, R. B. Macgregor, Jr. and C. E. Pearson (2014). "TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins." J Biol Chem **289**(8): 4653-4659.
Zaug, A. J., E. R. Podell and T. R. Cech (2005). "Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro." Proc Natl Acad Sci U S A **102**(31): 10864-10869.
Zhang, Q. S., L. Manche, R. M. Xu and A. R. Krainer (2006). "hnRNP A1 associates with telomere ends and stimulates telomerase activity." RNA **12**(6): 1116-1128.
Zhang, R., M. V. Poustovoitov, X. Ye, H. A. Santos, W. Chen, S. M. Daganzo, J. P. Erzberger, I. G. Serebriiskii, A. A. Canutescu, R. L. Dunbrack, J. R. Pehrson, J. M. Berger, P. D. Kaufman and P. D. Adams (2005). "Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA." Dev Cell **8**(1): 19-30.
Zhang, Y., H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird and D. Reinberg (1999). "Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation." Genes Dev **13**(15): 1924-1935.
Zhang, Y., C. L. Smith, A. Saha, S. W. Grill, S. Mihardja, S. B. Smith, B. R. Cairns, C. L. Peterson and C. Bustamante (2006). "DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC." Mol Cell **24**(4): 559-568.
Zhou, R., J. Zhang, M. L. Bochman, V. A. Zakian and T. Ha (2014). "Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA." Elife **3**: e02190.
Zhu, Y., M. J. Rowley, G. Bohmdorfer and A. T. Wierzbicki (2013). "A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing." Mol Cell **49**(2): 298-309.